# A new family of high-resolution multivariate spectral estimators

Mattia Zorzi



**Abstract**

In this paper, we extend the Beta divergence family to multivariate power spectral densities. Similarly to the scalar case, we show that it smoothly connects the multivariate *Kullback-Leibler* divergence with the multivariate *Itakura-Saito* distance. We successively study a spectrum approximation problem, based on the Beta divergence family, which is related to a multivariate extension of the THREE spectral estimation technique. It is then possible to characterize a family of solutions to the problem. An upper bound on the complexity of these solutions will also be provided. Simulations suggest that the most suitable solution of this family depends on the specific features required from the estimation problem.

**Index Terms**

Generalized covariance extension problem, Spectrum approximation problem, Structured covariance estimation problem, Beta divergence, Convex optimization

## I. Introduction

The recent development of THREE-like approaches to multivariate spectral estimation has triggered a renew interest for multivariate distance measures (or simply divergence indexes) among (power) spectral densities, [1]. In the THREE approach, the output covariance of a bank of filters is used to extract information on the input spectral density. More precisely,

This work was supported by the Italian Ministry for Education and Research (MIUR) under PRIN grant n. 20085FFJ2Z "New Algorithms and Applications of System Identification and Adaptive Control" and by University of Padova project "A Unifying Framework for Spectral Estimation and Matrix Completion: A New Paradigm for Identification, Estimation, and Signal Processing"

M. Zorzi is with the Dipartimento di Ingegneria dell'Informazione, Università di Padova, via Gradenigo 6/B, 35131 Padova, Italy, zorzimat@dei.unipd.it

the family of spectral densities matching the output covariance matrix is considered and a spectrum approximation problem, which "chooses" an estimate of the input spectral density in this family, is then employed. The choice criterium is based on finding the spectral density which minimizes a divergence index with respect to an *a priori* spectral density. Note that, the problem of parameterizing the family of feasible spectral densities may be viewed as a generalized covariance extension problem [2], [3], [4] [5], [6], [7]. The key feature for these estimators concerns the high resolution achievable in prescribed frequency bands, in particular with short data records. Significant applications to these methods can be found in $H_\infty$ robust control [8],[9], biomedical engineering [10], and modeling and identification [11], [12], [13].

The most delicate issue for this theory deals with the choice of the divergence index. In fact, the corresponding solution to the spectrum approximation problem (that heavily depends on the divergence index) must be computable and possibly with bounded *McMillan* degree. Accordingly, it is important to have many different indexes available in such a way to choose the most appropriate index in relation to the specific application. The THREE estimator, introduced by Byrnes Georgiou and Lindquist in [14], has been extended to the multichannel case by suggesting different multivariate divergence indexes, [15], [16], [17]. In particular, Georgiou introduced a multivariate version of the *Kullback-Leibler* divergence, [15], which has been frequently used within information theory, and a multivariate extension of the *Itakura-Saito* distance has been recently presented by Ferrante *et al.*, [17]. The latter metric has an interpretation in terms of relative entropy rate among processes. Finally, it is worth to note that the output covariance is not available in a THREE-like spectral estimation method. Indeed, we need to estimate it by using a collection of sample data generated by feeding the filters bank with the signal whose spectral density is to be estimated. Moreover, the family of spectral densities matching the estimated output covariance must be non-empty. This covariance estimation task is accomplished by solving a structured covariance estimation problem, [18], [19]. Therefore, a THREE-like spectral estimation procedure consists in solving a structured covariance estimation problem and then a spectrum approximation problem.

The main results of this paper are three. Firstly, we extend to the multivariate case the Beta divergence family (introduced for the scalar case in [20]) which smoothly connects the *Kullback-Leibler* divergence with the *Itakura-Saito* distance. It is worth mentioning that the Beta divergence family for scalar spectral densities has been widely used in many applications: Robust principal

component analysis and clustering [21], robust independent component analysis [22], and robust nonnegative matrix and tensor factorization [23], [24].

Secondly, we consider a spectrum approximation problem which employs the multivariate Beta divergence family. It turns out that it is possible to characterize a family of solutions to the problem with bounded *McMillan* degree. Moreover, its limit coincides to the solution obtained by using the *Kullback-Leibler* divergence.

Finally, we tackle the related structured covariance estimation problem which can be viewed as the static version of the previous spectrum approximation problem. Also in this case, a Beta matrix divergence family for covariance matrices, leading a family of solutions to the structured covariance estimation problem, may be introduced.

The paper is outlined as follows. Section II introduces THREE-like spectral estimation methods. Section III presents the new extension to the multivariate case of the Beta divergence family. In Section IV the corresponding spectrum approximation problem is introduced. More precisely, we derive the solution thanks to the means of the convex optimization. In Section V a non trivial existence result for the dual problem is established. Then, in Section VI a matricial *Newton* algorithm to efficiently solve the dual problem is presented. In Section VII some comparative examples are given: We test the different features of the found solutions. Section VIII is devoted to the estimation of structured covariance matrices by using the Beta matrix divergence family. Finally, in Section IX we propose an application to the estimation of multivariate spectral densities which employs the resulting THREE-like estimator.

## II. THREE-LIKE SPECTRAL ESTIMATION

Let us consider an unknown zero mean, $m$-dimensional, $\mathbb{R}^m$-valued, purely non-deterministic, stationary process $y = \{y_k;\ k \in \mathbb{Z}\}$ with spectral density $\Omega(\mathrm{e}^{j\vartheta})$ defined on the unit circle $\mathbb{T}$. Assume that the *a priori* information on $\Omega$ is given by a *prior* spectral density $\Psi \in \mathbb{S}_+^m(\mathbb{T})$. Here, $\mathbb{S}_+^m(\mathbb{T})$ denotes the family of bounded and coercive $\mathbb{R}^{m\times m}$-valued spectral density functions on $\mathbb{T}$. Then, a finite-length data $y_1 \dots y_N$ generated by $y$ is observed. We want to find an estimate $\Phi \in \mathbb{S}_+^m(\mathbb{T})$ of $\Omega$ by exploiting $\Psi$ and $y_1 \dots y_N$. This spectral estimation task is accomplished by employing a THREE-like approach which hinges on the following four elements:

1) A *prior* spectral density $\Psi \in \mathbb{S}_+^m(\mathbb{T})$;

2) A rational filter to process the data

$$G(z) = (zI - A)^{-1}B, \tag{1}$$

where $A \in \mathbb{R}^{n\times n}$ is a stability matrix, $B \in \mathbb{R}^{n\times m}$ is full rank with $n \geq m$, and $(A, B)$ is a reachable pair;

3) An estimate $\hat{\Sigma}$, based on the data $y_1 \dots y_N$, of the steady state covariance $\Sigma = \Sigma^T > 0$ of the state $x_k$ of the filter

$$x_{k+1} = Ax_k + By_k; \tag{2}$$

4) A divergence index $\mathcal{S}$ between two spectral densities.

According to the THREE-like approach, an estimate $\Phi \in \mathbb{S}_+^m(\mathbb{T})$ of $\Omega$ is given by solving the problem[1]:

$$\begin{aligned} &\text{minimize } \ \mathcal{S}(\Phi\|\Psi) \ \text{ over the set} \\ &\qquad \left\{ \Phi \in \mathbb{S}_+^m(\mathbb{T}) \mid \int G\Phi G^* = \hat{\Sigma} \right\}. \end{aligned} \tag{3}$$

Note that $\Psi$ is generally not consistent with $\hat{\Sigma}$, i.e. $\int G\Psi G^* \neq \hat{\Sigma}$. Hence, we have a spectrum approximation problem. The parametrization of all spectral densities satisfying constraint in (3) may be viewed as a generalized moment problem. For instance, the covariance extension problem may be recovered by setting

$$G(z) = \begin{bmatrix} z^{-n}I_m & \dots & z^{-1}I_m \end{bmatrix}^T. \tag{4}$$

In this case, the state covariance has a block *Toeplitz* structure:

$$\Sigma = \begin{bmatrix} \Sigma_0 & \Sigma_1 & \Sigma_2 & \dots & \Sigma_{n-1} \\ \Sigma_1^T & \Sigma_0 & \Sigma_1 & \ddots & \Sigma_{n-2} \\ \Sigma_2^T & \ddots & \ddots & \ddots & \ddots \\ \Sigma_{n-1}^T & \Sigma_{n-2}^T & \ddots & \ddots & \Sigma_0 \end{bmatrix}, \quad \Sigma_l = \mathbb{E}[y_k y_{k+l}^T]. \tag{5}$$

[1]Here and throughout the paper, integration, when not otherwise specified, is on the unit circle with respect to the normalized Lebesgue measure. Moreover, a star denotes transposition plus conjugation.

### A. Feasibility of the problem

The first issue arising with the previous spectrum approximation problem concerns its feasibility, i.e. the existence of $\Phi \in \mathbb{S}_+^m(\mathbb{T})$ satisfying the constraint in (3) for a given $\hat{\Sigma}$. To deal with this issue, we first introduce some notation: $\mathcal{Q}_n \subset \mathbb{R}^{n\times n}$ denotes the $n(n+1)/2$-dimensional real vector space of $n$-dimensional symmetric matrices and $\mathcal{Q}_{n,+}$ denotes the corresponding cone of positive definite matrices. We denote as $\mathcal{V}(\mathbb{S}_+^m)$ the linear space generated by $\mathbb{S}_+^m(\mathbb{T})$. Finally, we introduce the linear operator

$$\begin{aligned} \Gamma \;\; : \;\; & \mathcal{V}(\mathbb{S}_+^m) \to \mathcal{Q}_n \\ & \Phi \mapsto \int G\Phi G^*. \end{aligned} \tag{6}$$

In [25] (see also [16]), it was shown that a matrix $P \in \mathcal{Q}_n$ belongs to the range of $\Gamma$, denoted by $\mathrm{Range}\,\Gamma$, if and only if there exists $H \in \mathbb{R}^{m\times n}$ such that

$$P - APA^T = BH + H^T B^T. \tag{7}$$

An equivalent condition, [18], is that the kernel of the linear operator

$$\begin{aligned} V \;\; : \;\; & \mathcal{Q}_n \to \mathcal{Q}_n \\ & Q \mapsto \Pi_B^{\perp}(Q - AQA^T)\Pi_B^{\perp} \end{aligned} \tag{8}$$

contains $P$, namely $V(P) = 0$. Here, $\Pi_B^{\perp} := I - B(B^TB)^{-1}B^T$. It turns out that the spectrum approximation problem is feasible if and only if $\hat{\Sigma} \in \mathrm{Range}\,\Gamma \cap \mathcal{Q}_{n,+}$, [25],[16]. Let $x_1 \dots x_N$ be the output data generated by feeding the filters bank with the finite-length data $y_1 \dots y_N$. An estimate of $\Sigma$ is therefore given by the sample covariance matrix $\hat{\Sigma}_C := \frac{1}{N}\sum_{k=1}^N x_k x_k^T$ which is normally positive definite. It may not, however, belong to $\mathrm{Range}\,\Gamma$. Accordingly, we need to compute a new estimate $\hat{\Sigma} \in \mathrm{Range}\,\Gamma$ which is positive definite and "close" to the estimate $\hat{\Sigma}_C$. Hence, we have to solve a structured covariance estimation problem which lead us to consider the following convex optimization task.

*Problem 1:* Given $\hat{\Sigma}_C > 0$,

$$\begin{aligned} &\text{minimize } \; \mathcal{D}(P\|\hat{\Sigma}_C) \;\; \text{over the set} \\ &\qquad \{P \in \mathcal{Q}_{n,+} \mid V(P) = 0\}. \end{aligned} \tag{9}$$

Here, $\mathcal{D}$ is a suitable divergence index among (positive definite) covariance matrices. Problem 1 can be efficiently solved, [18], by considering the *information divergence* among two Gaussian densities with covariance $P$ and $Q$, respectively, [26]:

$$\mathcal{D}_{\mathrm{I}}(P\|Q) := \frac{1}{2}\,\mathrm{tr}[\log(Q) - \log(P) + PQ^{-1} - I]. \tag{10}$$

Another approach characterizes $\Sigma$ in terms of the filter parameters and the sequence of the covariance lags of $y$, [19]. Once we have $\hat{\Sigma}$ in such a way that the spectrum approximation problem is feasible, we can replace $G$ with $\overline{G} = \hat{\Sigma}^{-\frac{1}{2}}G$ and $(A, B)$ with $(\overline{A} = \hat{\Sigma}^{-\frac{1}{2}}A\hat{\Sigma}^{\frac{1}{2}}, \overline{B} = \hat{\Sigma}^{-\frac{1}{2}}B)$. Thus, the constraint may be rewritten as $\int \bar{G}\Phi\bar{G}^* = I$. Accordingly, from now on we assume that the spectrum approximation problem in (3) is feasible and we consider the following equivalent formulation.

*Problem 2:* Given $\Psi \in \mathbb{S}_+^m(\mathbb{T})$ and $G(z) = (zI - A)^{-1}B$ such that $I \in \mathrm{Range}\,\Gamma$,

$$\begin{aligned} &\text{minimize } \mathcal{S}(\Phi\|\Psi) \text{ over the set} \\ &\qquad \left\{\Phi \in \mathbb{S}_+^m(\mathbb{T}) \mid \int G\Phi G^* = I\right\}. \end{aligned} \tag{11}$$

### *B. Choice of the divergence index*

A divergence index among spectral densities in $\mathbb{S}_+^m(\mathbb{T})$ must satisfy the following basic property for all $\Phi, \Psi \in \mathbb{S}_+^m(\mathbb{T})$:

$$\begin{aligned} &\mathcal{S}(\Phi\|\Psi) \geq 0 \\ &\mathcal{S}(\Phi\|\Psi) = 0 \text{ if and only if } \Phi = \Psi. \end{aligned} \tag{12}$$

Moreover, the corresponding Problem 2 should lead to a computable solution, by typically solving the dual optimization problem. In [15], a *Kullback-Leibler* divergence for multivariate spectral densities with the same trace of the zeroth-moment has been introduced

$$\mathcal{S}_{KL0}(\Phi\|\Psi) = \int \mathrm{tr}[\Phi(\log(\Phi) - \log(\Psi))] \tag{13}$$

where $\log(\cdot)$, whose definition will be given in Section III-B, is the matrix logarithm. This divergence is inspired by the *Umegaki-von Neumann*'s relative entropy [27] of statistical quantum mechanics. Moreover, (13) may be readily extended to the general case, see [28] for the scalar case,

$$\mathcal{S}_{\mathrm{KL}}(\Phi\|\Psi) = \int \mathrm{tr}[\Phi(\log(\Phi) - \log(\Psi)) - \Phi + \Psi] \tag{14}$$

and $\mathcal{S}_{KL0}(\Phi\|\Psi) = \mathcal{S}_{\mathrm{KL}}(\Phi\|\Psi)$ when $\int \mathrm{tr}\,\Phi = \int \mathrm{tr}\,\Psi$. The resulting solution to the spectrum approximation problem is, however, not rational, even when $\Psi = I$. On the contrary, by considering the multivariate extension of the *Itakura-Saito* distance

$$\mathcal{S}_{\mathrm{IS}}(\Phi\|\Psi) = \int \mathrm{tr}[\log(\Psi) - \log(\Phi) + \Phi\Psi^{-1} - I], \tag{15}$$

a rational solution is obtained, [17]. We will show in the following Section that the divergence indexes (14) and (15) belongs to the same multivariate Beta divergence family. Moreover, this family leads to a family of solutions to the spectrum approximation problem.

Observe that, it is also possible to rewrite Problem 2 by considering $\mathcal{S}_{\mathrm{KL}}(\Psi\|\Phi)$. The resulting solution is, however, only computable when $y$ is a scalar process [14],[29], or $\Psi = I$, [5], [30], [15]. Finally, we mention that there exists another multivariate distance, called *Hellinger* distance, which gives a rational solution to Problem 2, [16].

## III. Beta divergence family for spectral densities

In this section we extend the notion of Beta divergence (family) for scalar spectral densities, firstly introduced in [20] and [22], to the multivariate case.

### A. Scalar case

We recall the definition of the scalar Beta divergence by adopting the same notation employed in [28]. First of all, we need to introduce the following function

$$\begin{aligned} \log_c : \;& \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R} \\ & (x,y) \mapsto \begin{cases} \frac{1}{1-c}\left[\left(\frac{x}{y}\right)^{1-c} - 1\right], & c \in \mathbb{R}\setminus\{1\} \\ \log(x) - \log(y), & c = 1 \end{cases} \end{aligned} \tag{16}$$

which is referred to as *generalized logarithm discrepancy* function throughout the paper. Notice that $\log_c$ is a continuous function of real variable $c$ and $\log_c(x,y) = 0$ if and only if $x = y$. The (asymmetric) Beta divergence between two scalar spectral densities $\Phi, \Psi \in \mathbb{S}^1_+(\mathbb{T})$ is defined by

$$\begin{aligned} \mathcal{S}_\beta(\Phi\|\Psi) &:= -\frac{1}{\beta}\int [\Phi^\beta \log_{\frac{1}{\beta}}\left(\Psi^\beta, \Phi^\beta\right) + \Phi^\beta - \Psi^\beta] \\ &= \int [\frac{1}{\beta-1}(\Phi^\beta - \Phi\Psi^{\beta-1}) - \frac{1}{\beta}(\Phi^\beta - \Psi^\beta)] \end{aligned} \tag{17}$$

where the parameter $\beta$ is a real number. For $\beta = 0$ and $\beta = 1$, it is defined by continuity in the following way

$$\begin{aligned} \lim_{\beta \to 0} \mathcal{S}_\beta(\Phi \| \Psi) &= \mathcal{S}_{\mathrm{IS}}(\Phi \| \Psi) \\ \lim_{\beta \to 1} \mathcal{S}_\beta(\Phi \| \Psi) &= \mathcal{S}_{\mathrm{KL}}(\Phi \| \Psi), \end{aligned} \tag{18}$$

where $\mathcal{S}_{\mathrm{IS}}$ and $\mathcal{S}_{\mathrm{KL}}$ are the scalar versions of (15) and (14), respectively. Moreover, the Beta divergence is a continuous function of real variable $\beta$ in the whole range including singularities. Thus, it smoothly connects the *Itakura-Saito* distance with the *Kullback-Leibler* divergence. Since $\mathcal{S}_\beta$ is a divergence index, property (12) is satisfied. Finally, $\mathcal{S}_\beta$ is always strictly convex in the first argument, but is often not in the second argument.

### *B. Multivariate case*

Likewise to the scalar case, we start by introducing the *generalized multivariate logarithm discrepancy*. To this aim, recall that the exponentiation of a positive definite matrix $X$ to an arbitrary real number c, is defined as $X^c := U\mathrm{diag}(d_1^c, \ldots, d_m^c)U^T$ where $X := U\mathrm{diag}(d_1, \ldots, d_m)U^T$ is the usual spectral decomposition with $U$ orthogonal, i.e. $UU^T = I$, and $\mathrm{diag}(d_1, \ldots, d_m) > 0$ diagonal matrix.[2] We are now ready to extend the definition of *generalized logarithm discrepancy* to the multivariate case

$$\begin{aligned} \log_c : \quad & \mathcal{Q}_{m,+} \times \mathcal{Q}_{m,+} \to \mathcal{Q}_m \\ & (X, Y) \mapsto \begin{cases} \frac{1}{1-c}(X^{1-c}Y^{c-1} - I), & c \in \mathbb{R} \setminus \{1\} \\ \log(X) - \log(Y), & c = 1 \end{cases} \end{aligned} \tag{19}$$

where $\log(X) = U\mathrm{diag}(\log(d_1), \ldots, \log(d_m))U^T$ is the matrix logarithm of $X$.

*Proposition 3.1:* The *generalized multivariate logarithm discrepancy* is a continuous function of real variable $c$ in the whole range. Moreover, $\log_c(X, Y) = 0$ if and only if $X = Y$.

*Proof:* By definition $X^{1-c}$ and $Y^{c-1}$ are continuous function of real variable $c$. Thus, the function $\log_c(X, Y)$ of real variable $c$ is continuous in $\mathbb{R} \setminus \{1\}$. It remains to prove that $\log_c$ is continuous in $c = 1$. This is equivalent to show that $\lim_{c \to 1} \log_c(X, Y) = \log(X) - \log(Y)$. Let $X = U\mathrm{diag}(d_1, \ldots, d_m)U^T$, then

$$\frac{1}{1-c}(X^{1-c} - I) = U\mathrm{diag}(\frac{d_1^{1-c} - 1}{1-c}, \ldots, \frac{d_m^{1-c} - 1}{1-c})U^T. \tag{20}$$

[2]It is also possible to take the exponentiation of positive semidefinite matrices when $c \neq 0$.

Taking the limit for $c \to 1$, we get

$$\begin{aligned}
&\lim_{c\to 1} \frac{1}{1-c}(X^{1-c} - I) \\
&\quad = U\mathrm{diag}(\lim_{c\to 1} \frac{d_1^{1-c}-1}{1-c}, \ldots, \lim_{c\to 1} \frac{d_m^{1-c}-1}{1-c})U^T \\
&\quad = U\mathrm{diag}(\log(d_1), \ldots, \log(d_m))U^T = \log(X). \qquad (21)
\end{aligned}$$

Accordingly,

$$\begin{aligned}
&\lim_{c\to 1} \log_c(X,Y) \\
&\quad = \lim_{c\to 1}[\frac{1}{1-c}(X^{1-c} - I) - \frac{1}{1-c}(Y^{1-c} - I)]Y^{c-1} \\
&\quad = \lim_{c\to 1}[\frac{1}{1-c}(X^{1-c} - I)] - \lim_{c\to 1}[\frac{1}{1-c}(Y^{1-c} - I)] \\
&\quad = \log(X) - \log(Y) \qquad (22)
\end{aligned}$$

which proves that $\log_c$ is continuous in $c = 1$. Concerning the last statement, it is straightforward that $X = Y$ implies $\log_c(X,Y) = 0$. On the contrary, $\log_c(X,Y) = 0$, with $c \neq 1$, implies $X^{1-c}Y^{c-1} = I$ which is equivalent to $X^{1-c} = Y^{1-c}$, since $X, Y \in \mathcal{Q}_{m,+}$. Thus, $X = Y$. We get the same conclusion for $c = 1$ by exploiting similar argumentations. ■

The exponentiation of a spectral density $\Phi(\mathrm{e}^{j\vartheta}) \in \mathbb{S}_+^m(\mathbb{T})$ to an arbitrary real number $c$ is punctually defined by exploiting the previous spectral decomposition:

$$\Phi(\mathrm{e}^{j\vartheta})^c = U(\mathrm{e}^{j\vartheta})\mathrm{diag}(d_1(\mathrm{e}^{j\vartheta})^c, \ldots, d_m(\mathrm{e}^{j\vartheta})^c)U(\mathrm{e}^{j\vartheta})^T$$

where $\Phi(\mathrm{e}^{j\vartheta}) = U(\mathrm{e}^{j\vartheta})\mathrm{diag}(d_1(\mathrm{e}^{j\vartheta}), \ldots, d_m(\mathrm{e}^{j\vartheta}))U(\mathrm{e}^{j\vartheta})^T$ with $U(\mathrm{e}^{j\vartheta}) \in \mathrm{L}_\infty^{m\times m}(\mathbb{T})$ such that $U(\mathrm{e}^{j\vartheta})U(\mathrm{e}^{j\vartheta})^T = I$. Observe that $\Phi^c$ belongs to $\mathbb{S}_+^m(\mathbb{T})$. We are now ready to introduce the multivariate (asymmetric) Beta divergence among $\Phi, \Psi \in \mathbb{S}_+^m(\mathbb{T})$:

$$\begin{aligned}
\mathcal{S}_\beta(\Phi\|\Psi) &:= -\frac{1}{\beta}\int \mathrm{tr}[\Phi^\beta \log_{\frac{1}{\beta}}\left(\Psi^\beta, \Phi^\beta\right) + \Phi^\beta - \Psi^\beta] \\
&= \int \mathrm{tr}[\frac{1}{\beta-1}(\Phi^\beta - \Phi\Psi^{\beta-1}) - \frac{1}{\beta}(\Phi^\beta - \Psi^\beta)] \qquad (23)
\end{aligned}$$

where $\beta \in \mathbb{R} \setminus \{0,1\}$. Similarly to the scalar case, we can extend by continuity the definition of Beta divergence for $\beta = 0$ and $\beta = 1$.

*Proposition 3.2:* The following limits hold:

$$\begin{aligned}
\lim_{\beta\to 0} \mathcal{S}_\beta(\Phi\|\Psi) &= \mathcal{S}_{\mathrm{IS}}(\Phi\|\Psi) \\
\lim_{\beta\to 1} \mathcal{S}_\beta(\Phi\|\Psi) &= \mathcal{S}_{\mathrm{KL}}(\Phi\|\Psi). \qquad (24)
\end{aligned}$$

*Proof:* Since $\Phi$ and $\Psi$ belong to $\mathbb{S}_+^m(\mathbb{T})$, i.e. $\Phi$ and $\Psi$ are coercive and bounded, it is possible to show by standard argumentations that the integrand function of (23) uniformly converges on $\mathbb{T}$ for $\beta \to 0$ and $\beta \to 1$. Hence, it is allowed to pass the limits, for $\beta \to 0$ and $\beta \to 1$, under the integral sign. Taking into account the first limit, we get

$$\begin{aligned}
&\lim_{\beta\to 0} \mathcal{S}_\beta(\Phi\|\Psi) \\
&= \lim_{\beta\to 0} \int \mathrm{tr}[\frac{1}{\beta-1}(\Phi^\beta - \Phi\Psi^{\beta-1}) - \frac{1}{\beta}(\Phi^\beta - \Psi^\beta)] \\
&= \int \mathrm{tr}\left\{-I + \Phi\Psi^{-1} - \lim_{\beta\to 0}\frac{1}{\beta}[(\Phi^\beta - I) - (\Psi^\beta - I)]\right\} \\
&= \int \mathrm{tr}[-I + \Phi\Psi^{-1} - \log(\Phi) + \log(\Psi)] \\
&= \mathcal{S}_{\mathrm{IS}}(\Phi\|\Psi) \qquad (25)
\end{aligned}$$

where we exploited (21). For the second limit, we obtain

$$\begin{aligned}
&\lim_{\beta\to 1} \mathcal{S}_\beta(\Phi\|\Psi) \\
&= \lim_{\beta\to 1}\left\{-\frac{1}{\beta}\int \mathrm{tr}[\Phi^\beta \log_{\frac{1}{\beta}}\left(\Psi^\beta, \Phi^\beta\right) + \Phi^\beta - \Psi^\beta]\right\} \\
&= -\int \mathrm{tr}[\Phi \lim_{\beta\to 1} \log_{\frac{1}{\beta}}\left(\Psi^\beta, \Phi^\beta\right) + \Phi - \Psi] \\
&= -\int \mathrm{tr}[\Phi \lim_{\beta\to 1} \log_{2-\beta}\left(\Psi, \Phi\right) + \Phi - \Psi] \\
&= \int \mathrm{tr}[\Phi\left(\log(\Phi) - \log(\Psi)\right) + \Psi - \Phi] \\
&= \mathcal{S}_{\mathrm{KL}}(\Phi\|\Psi) \qquad (26)
\end{aligned}$$

where we exploited (22). ■

In view of Proposition 3.1 and Proposition 3.2, we conclude that the multivariate Beta divergence is a continuous function of real variable $\beta$ in the whole range including singularities and it smoothly connects the multivariate *Itakura-Saito* distance with the multivariate *Kullback-Leibler* divergence.

*Remark 3.1:* For $\beta = 2$, the Beta divergence corresponds, up to a constant scalar factor, to the standard squared *Euclidean* distance ($L_2$-norm)

$$\mathcal{S}_{\mathrm{L2}}(\Phi\|\Psi) = \int \langle \Phi - \Psi, \Phi - \Psi \rangle \qquad (27)$$

where $\langle X, Y\rangle = \mathrm{tr}(XY)$ is the usual scalar product in $\mathcal{Q}_m$.

Finally, we show that the multivariate Beta divergence satisfies condition (12).

*Proposition 3.3:* Given $\Phi, \Psi \in \mathbb{S}_+^m(\mathbb{T})$, the following facts hold:

1) $\mathcal{S}_\beta(\cdot\|\Psi)$ is strictly convex over $\mathbb{S}_+^m(\mathbb{T})$,

2) $\mathcal{S}_\beta(\Phi\|\Psi) \geq 0$ and equality holds if and only if $\Psi = \Phi$.

*Proof:* In order to prove the statements, we need of the following first variations of the maps $X \mapsto \mathrm{tr}(X^c)$ and $X \mapsto \mathrm{tr}(X^c Y)$, respectively (further details may be found in Appendix A):

$$\begin{aligned}\delta(\mathrm{tr}[X^c]; \delta X) &= c\,\mathrm{tr}[X^{c-1}\delta X]\\ \delta(\mathrm{tr}(X^c Y); \delta X) &= \mathrm{tr}[O_{X,c}(\delta X)Y],\end{aligned} \tag{28}$$

where $Y \in \mathcal{Q}_m$ and the map $O_{X,c}$ is defined in (78).

1) The first variation of $\mathcal{S}_\beta(\Phi\|\Psi)$, with respect to $\Phi$, in direction $\delta\Phi \in \mathrm{L}_\infty^{m\times m}(\mathbb{T})$ is

$$\delta(\mathcal{S}_\beta(\Phi\|\Psi); \delta\Phi) = \frac{1}{\beta-1}\int_0^{2\pi} \mathrm{tr}[(\Phi^{\beta-1} - \Psi^{\beta-1})\delta\Phi]\frac{\mathrm{d}\vartheta}{2\pi}. \tag{29}$$

The second variation in direction $\delta\Phi$ is

$$\begin{aligned}\delta^2(\mathcal{S}_\beta(\Phi\|\Psi); \delta\Phi) &= \int_0^{2\pi} \mathrm{tr}[O_{\Phi,\beta-1}(\delta\Phi)\delta\Phi]\frac{\mathrm{d}\vartheta}{2\pi}\\ &= \int_0^{2\pi} \mathrm{tr}\left[\int_0^1 \Phi^{(\beta-1)(1-\tau)}\int_0^\infty (\Phi+tI)^{-1}\delta\Phi\right.\\ &\qquad \left.\times\,(\Phi+tI)^{-1}\mathrm{d}t\Phi^{(\beta-1)\tau}\mathrm{d}\tau\delta\Phi\right]\frac{\mathrm{d}\vartheta}{2\pi}\\ &= \int_0^{2\pi}\int_0^1\int_0^\infty \mathrm{tr}[\Phi^{(\beta-1)(1-\tau)}(\Phi+tI)^{-1}\delta\Phi\\ &\qquad \times(\Phi+tI)^{-1}\Phi^{(\beta-1)\tau}\delta\Phi]\mathrm{d}t\mathrm{d}\tau\frac{\mathrm{d}\vartheta}{2\pi}.\end{aligned}$$

By the cyclic property of the trace and since $\Phi^{(\beta-1)\tau}$ and $(\Phi+tI)^{-1}$ commute, we get

$$\delta^2(\mathcal{S}_\beta(\Phi\|\Psi); \delta\Phi) = \int_0^{2\pi}\int_0^1\int_0^\infty f_{t,\tau}(\Phi, \delta\Phi)\mathrm{d}t\mathrm{d}\tau\frac{\mathrm{d}\vartheta}{2\pi} \tag{30}$$

where

$$\begin{aligned}f_{t,\tau}(X,\Delta) &= \mathrm{tr}[X^{\frac{(\beta-1)\tau}{2}}(X+tI)^{-\frac{1}{2}}\Delta(X+tI)^{-\frac{1}{2}}\\ &\quad \times X^{(\beta-1)(1-\tau)}(X+tI)^{-\frac{1}{2}}\Delta(X+tI)^{-\frac{1}{2}}X^{\frac{(\beta-1)\tau}{2}}]\end{aligned} \tag{31}$$

with $X \in \mathcal{Q}_{m,+}$, $\Delta \in \mathcal{Q}_m$, $t \in [0,\infty)$ and $\tau \in [0,1]$. Thus, $f_{t,\tau}(X,\Delta) \geq 0$ and $f_{t,\tau}(X,\Delta) = 0$ if and only if $\Delta = 0$. We conclude that integral (30), i.e. the second variation of $\mathcal{S}_\beta(\cdot\|\Psi)$, is positive for $\delta\Phi \neq 0$. Accordingly, $\mathcal{S}_\beta(\cdot\|\Psi)$ is strictly convex over the convex set $\mathbb{S}^m_+(\mathbb{T})$.
2) As a consequence of the previous statement, the minimum point is unique and it is given by annihilating (29) for each $\delta\Phi \in \mathrm{L}^{m\times m}_\infty(\mathbb{T})$. Since $\Phi^{\beta-1} - \Psi^{\beta-1} \in \mathrm{L}^{m\times m}_\infty(\mathbb{T})$, it follows that the minimum point satisfies the condition $\Phi^{\beta-1} = \Psi^{\beta-1}$. Accordingly, $\Phi = \Psi$. Finally it is sufficient to observe that $\mathcal{S}_\beta(\Psi\|\Psi) = 0$. ■

Note that $\mathcal{S}_\beta(\Phi\|\cdot)$ is not convex on $\mathbb{S}^m_+(\mathbb{T})$ (not even in the scalar case).

## IV. Spectrum approximation problem

Since the Beta divergence is well-defined for $\beta \in \mathbb{R}$, we choose $\beta = -\frac{1}{\nu} + 1$ with $\nu \in \mathbb{Z} \setminus \{0\}$. As we will see, this choice guarantees that the corresponding solution to Problem 2 (which is feasible) is rational for a suitable choice of $\Psi$. In order to simplify the notation we define $\mathcal{S}_\nu(\Phi\|\Psi) := \mathcal{S}_\beta(\Phi\|\Psi)$ with $\beta = -\frac{1}{\nu} + 1$. We have to minimize $\mathcal{S}_\nu(\Phi\|\Psi)$ over $\left\{\Phi \in \mathbb{S}^m_+(\mathbb{T}) \mid \int G\Phi G^* = I\right\}$. Since it is a constrained convex optimization problem, we consider the corresponding *Lagrange* functional

$$\begin{aligned} &L_\nu(\Phi,\Lambda) \\ &\quad = \mathcal{S}_\nu(\Phi\|\Psi) + \frac{\nu}{1-\nu}\int \mathrm{tr}(\Psi^{\frac{\nu-1}{\nu}}) + \left\langle \int G\Phi G^* - I, \Lambda \right\rangle \\ &\quad = \int \mathrm{tr}\left[-\nu(\Phi^{\frac{\nu-1}{\nu}} - \Phi\Psi^{-\frac{1}{\nu}}) + \frac{\nu}{1-\nu}\Phi^{\frac{\nu-1}{\nu}} + G^*\Lambda G\Phi\right] - \mathrm{tr}(\Lambda) \end{aligned} \tag{32}$$

where we exploited the fact that the term $\int \mathrm{tr}(\Psi^{\frac{\nu-1}{\nu}})$ plays no role in the optimization problem. Note that, the *Lagrange* multiplier $\Lambda \in \mathcal{Q}_n$ can be uniquely decomposed as $\Lambda = \Lambda_\Gamma + \Lambda_\perp$ where $\Lambda_\Gamma \in \mathrm{Range}\ \Gamma$, $\Lambda_\perp \in [\mathrm{Range}\ \Gamma]^\perp$. Since $\Lambda_\perp$ is such that $G^*(\mathrm{e}^{j\vartheta})\Lambda_\perp G(\mathrm{e}^{j\vartheta}) \equiv 0$ and $\mathrm{tr}(\Lambda_\perp) = \langle \Lambda_\perp, I\rangle = 0$ (see [31, Section III]), it does not affect the *Lagrangian*, i.e. $L_\nu(\Phi,\Lambda) = L_\nu(\Phi,\Lambda_\Gamma)$. Accordingly we can impose from now on that $\Lambda \in \mathrm{Range}\ \Gamma$.

Consider now the unconstrained minimization problem $\min_\Phi \left\{L_\nu(\Phi,\Lambda) \mid \Phi \in \mathbb{S}^m_+(\mathbb{T})\right\}$. Since $L_\nu(\cdot,\Lambda)$ is strictly convex over $\mathbb{S}^m_+(\mathbb{T})$, its unique minimum point $\Phi_\nu$ is given by annihilating its first variation in each direction $\delta\Phi \in \mathrm{L}^{m\times m}_\infty(\mathbb{T})$:

$$\delta L_\nu(\Phi,\Lambda;\delta\Phi) = \int \mathrm{tr}\left[\left(\nu(\Psi^{-\frac{1}{\nu}} - \Phi^{-\frac{1}{\nu}}) + G^*\Lambda G\right)\delta\Phi\right] \tag{33}$$

where we exploited (28). Note that, $\nu(\Psi^{-\frac{1}{\nu}} - \Phi^{-\frac{1}{\nu}}) + G^*\Lambda G \in \mathrm{L}_\infty^{m\times m}(\mathbb{T})$. Thus, (33) is zero $\forall \Phi \in \mathrm{L}_\infty^{m\times m}(\mathbb{T})$ if and only if

$$\Phi^{-\frac{1}{\nu}} = \Psi^{-\frac{1}{\nu}} + \frac{1}{\nu} G^*\Lambda G. \tag{34}$$

Since $\Phi^{-\frac{1}{\nu}} \in \mathbb{S}_+^m(\mathbb{T})$, the set of the admissible *Lagrange* multipliers is

$$\mathcal{L}_\nu := \left\{ \Lambda \in \mathcal{Q}_n \mid \Psi^{-\frac{1}{\nu}} + \frac{1}{\nu} G^*\Lambda G > 0 \text{ on } \mathbb{T} \right\}. \tag{35}$$

Therefore, the natural set for $\Lambda$ is

$$\mathcal{L}_\nu^\Gamma = \mathcal{L}_\nu \cap \mathrm{Range}\ \Gamma. \tag{36}$$

In conclusion, the unique minimum point of the *Lagrange* functional has the form

$$\Phi_\nu(\Lambda) := [\Psi^{-\frac{1}{\nu}} + \frac{1}{\nu} G^*\Lambda G]^{-\nu}. \tag{37}$$

Assuming that $\Psi^{\frac{1}{\nu}}$ is rational in $\mathrm{e}^{j\vartheta}$, there always exists a unique (up to a right-multiplication by a constant orthogonal matrix) stable and minimum phase spectral factor $W$ such that $\Psi(\mathrm{e}^{j\vartheta})^{\frac{1}{\nu}} = W(\mathrm{e}^{j\vartheta})W(\mathrm{e}^{j\vartheta})^*$. By defining $G_1(\mathrm{e}^{j\vartheta}) = \frac{1}{\sqrt{\nu}} G(\mathrm{e}^{j\vartheta})W(\mathrm{e}^{j\vartheta})$, we obtain an equivalent form of (37):

$$\Phi_\nu(\Lambda) = [W(I + G_1^*\Lambda G_1)^{-1}W^*]^\nu. \tag{38}$$

*Corollary 4.1:* Let $\Psi$ be a constant *prior*. Then, $\Phi_\nu$ is rational in $\mathrm{e}^{j\vartheta}$ with *McMillan* degree less than or equal to $2n|\nu|$. Moreover, among all the spectral densities $\Phi_\nu$ with $\nu \in \mathbb{Z} \setminus \{0\}$:

1) The spectral densities with the smallest upper bound on the McMillan degree correspond to the *Itakura-Saito* and the squared *Euclidean* distance
2) As $\nu \to \pm\infty$, $\Phi_\nu$ tends to the spectral density corresponding to the *Kullback-Leibler* divergence.

*Proof:* In view of (37) and (38), $\deg[\Phi_\nu] \leq |\nu|(\deg[\Psi^{\frac{1}{\nu}}] + 2n) = 2n|\nu|$ where $n$ is the *McMillan* degree of $G(z)$.

1) Since $\nu \in \mathbb{Z} \setminus \{0\}$, the spectral densities with the smallest upper bound on the *McMillan* degree are attained for $\nu = \pm 1$, i.e. $\beta = 0$ and $\beta = 2$, which are the optimal forms related to $\mathcal{S}_{\mathrm{IS}}(\Phi\|\Psi)$ and $\mathcal{S}_{\mathrm{L2}}(\Phi\|\Psi)$, respectively. Note that, $\Phi_1(\Lambda) = [\Psi^{-1} + G^*\Lambda G]^{-1}$, which is the same optimal form found in [17] for the multivariate *Itakura-Saito* distance, and $\Phi_{-1}(\Lambda) = \Psi - G^*\Lambda G$.

2) Firstly, it is possible to show that the optimal form obtained by using the *Kullback-Leibler*

divergence is $\Phi_{\mathrm{KL}}(\Lambda) = \mathrm{e}^{\log(\Psi) - G^*\Lambda G}$ which is a straightforward generalization of the optimal form for $\mathcal{S}_{\mathrm{KL}0}$ presented in [15]. We want to show that $\Phi_\nu \to \Phi_{\mathrm{KL}}$ as $\nu \to \pm\infty$. Let us consider the function $F(\lambda) := \log(\Psi^{-\lambda} + \lambda G^*\Lambda G)$ with $\lambda \in \mathbb{R}$ such that $\Psi^{-\lambda} + \lambda G^*\Lambda G > 0$ on $\mathbb{T}$. Its first order *Taylor* expansion with respect to $\lambda = 0$ is $\Psi^{-\lambda} + \lambda G^*\Lambda G - I$. Accordingly,

$$\begin{aligned}
&\lim_{\nu\to\pm\infty} \nu \log(\Psi^{-\frac{1}{\nu}} + \frac{1}{\nu} G^*\Lambda G) \\
&= \lim_{\nu\to\pm\infty} \frac{\Psi^{-\frac{1}{\nu}} - I}{\nu^{-1}} + G^*\Lambda G \\
&= -\log(\Psi) + G^*\Lambda G
\end{aligned} \tag{39}$$

where we exploited (21) and the previous *Taylor* expansion. Finally,

$$\begin{aligned}
\lim_{\nu\to\pm\infty} \Phi_\nu(\Lambda) &= \lim_{\nu\to\pm\infty} \mathrm{e}^{\log[(\Psi^{-\frac{1}{\nu}} + \frac{1}{\nu} G^*\Lambda G)^{-\nu}]} \\
&= \lim_{\nu\to\pm\infty} \mathrm{e}^{-\nu \log(\Psi^{-\frac{1}{\nu}} + \frac{1}{\nu} G^*\Lambda G)} \\
&= \mathrm{e}^{-\lim_{\nu\to\pm\infty} \nu \log(\Psi^{-\frac{1}{\nu}} + \frac{1}{\nu} G^*\Lambda G)} \\
&= \mathrm{e}^{\log(\Psi) - G^*\Lambda G} = \Phi_{\mathrm{KL}}(\Lambda).
\end{aligned} \tag{40}$$

■

In this section we showed that $\Phi_\nu(\Lambda)$ is the unique minimum point of $L_\nu(\cdot, \Lambda)$, namely

$$L_\nu(\Phi_\nu(\Lambda), \Lambda) < L_\nu(\Phi, \Lambda), \quad \forall \Phi \in \mathbb{S}_+^m(\mathbb{T}) \text{ s.t. } \Phi \neq \Phi_\nu(\Lambda). \tag{41}$$

Hence, if we produce $\Lambda^\circ \in \mathcal{L}_\nu^\Gamma$ satisfying constraint in (11), inequality (41) implies

$$\mathcal{S}_\nu(\Phi_\nu(\Lambda^\circ) \| \Psi) < \mathcal{S}_\nu(\Phi \| \Psi), \quad \forall \Phi \in \mathbb{S}_+^m(\mathbb{T}) \text{ s.t. } \Phi \neq \Phi_\nu(\Lambda^\circ) \tag{42}$$

namely such a $\Phi_\nu(\Lambda^\circ)$ is the unique solution to Problem 2 with $\mathcal{S}_\nu$. The following step consists in showing the existence of such a $\Lambda^\circ$ by exploiting the duality theory.

## V. DUAL PROBLEM

Here, we do not deal with the case $\nu = 1$, since the existence of the solution to the dual problem was already showed in [17]. We start by considering the case $\nu \in \mathbb{N}_+ \setminus \{1\}$[3]. The dual

[3] $\mathbb{N}_+$ denotes the set of the positive natural numbers.

problem consists in maximizing the functional

$$\begin{aligned}\inf_{\Phi} L_\nu(\Phi,\Lambda) &= L_\nu(\Phi_\nu,\Lambda)\\ &= \frac{\nu}{1-\nu}\int \mathrm{tr}[(\Psi^{-\frac{1}{\nu}}+\frac{1}{\nu}G^*\Lambda G)^{1-\nu}]-\mathrm{tr}(\Lambda) \qquad (43)\end{aligned}$$

which is equivalent to minimize the following functional hereafter referred to as *dual functional*:

$$J_\nu(\Lambda) = -\frac{\nu}{1-\nu}\int \mathrm{tr}[(\Psi^{-\frac{1}{\nu}}+\frac{1}{\nu}G^*\Lambda G)^{1-\nu}]+\mathrm{tr}(\Lambda). \qquad (44)$$

*Theorem 5.1:* The dual functional $J_\nu$ belongs to $\mathcal{C}^\infty(\mathcal{L}_\nu^\Gamma)$ and it is strictly convex over $\mathcal{L}_\nu^\Gamma$.

*Proof:* In view of (28), the first variation of $J_\nu(\Lambda)$ in direction $\delta\Lambda_1 \in \mathcal{Q}_n$ is

$$\begin{aligned}&\delta J_\nu(\Lambda;\delta\Lambda_1)\\ &= -\int \mathrm{tr}[(\Psi^{-\frac{1}{\nu}}+\frac{1}{\nu}G^*\Lambda G)^{-\nu}G^*\delta\Lambda_1 G]+\mathrm{tr}(\delta\Lambda_1)\\ &= -\int \mathrm{tr}\left[\left(W(I+G_1^*\Lambda G_1)^{-1}W^*\right)^\nu G^*\delta\Lambda_1 G\right]+\mathrm{tr}(\delta\Lambda_1). \qquad (45)\end{aligned}$$

The linear form $\nabla J_{\nu,\Lambda}(\cdot) := \delta J_\nu(\Lambda;\cdot)$ is the *gradient* of $J_\nu$ at $\Lambda$. In order to prove that $J_\nu(\Lambda) \in \mathcal{C}^1(\mathcal{L}_\nu^\Gamma)$ we have to show that $\delta(J_\nu(\Lambda);\delta\Lambda_1)$, for any fixed $\delta\Lambda_1$, is continuous in $\Lambda$. To this aim, consider a sequence $M_n \in \mathrm{Range}\ \Gamma$ such that $M_n \to 0$ and define $Q_N(z) = W(z)(I + G_1(z)^*NG_1(z))^{-1}W(z)^*$ with $N \in \mathcal{Q}_n$. By Lemma 5.2 in [31] and since $W$ is bounded on $\mathbb{T}$, $Q_{\Lambda+M_n}$ converges uniformly to $Q_\Lambda$. Thus, applying elementwise the bounded convergence theorem, we obtain

$$\lim_{n\to\infty}\int GQ_{\Lambda+M_n}^\nu G^* = \int GQ_\Lambda^\nu G^*. \qquad (46)$$

Accordingly, $\delta(J_\nu(\Lambda);\delta\Lambda)$ is continuous, i.e. $J_\nu$ belongs to $\mathcal{C}^1(\mathcal{L}_\nu^\Gamma)$. In order to compute the second variation, consider the operator $\mathcal{I} : A \mapsto A^{-\nu}$. By applying the chain rule, we get

$$\delta(\mathcal{I}(A);\delta A) = -\sum_{l=1}^{\nu} A^{-l}\delta A A^{l-\nu-1}. \qquad (47)$$

Thus, for $\delta\Lambda_1, \delta\Lambda_2 \in \mathcal{Q}_n$ we have

$$\begin{aligned}&\delta^2 J_\nu(\Lambda;\delta\Lambda_1,\delta\Lambda_2)\\ &= \frac{1}{\nu}\sum_{l=1}^{\nu}\int \mathrm{tr}[Q_\Lambda^l G^*\delta\Lambda_2 G Q_\Lambda^{\nu+1-l}G^*\delta\Lambda_1 G]. \qquad (48)\end{aligned}$$

The bilinear form $\mathcal{H}_{\nu,\Lambda}(\cdot,\cdot) = \delta^2 J_\nu(\Lambda;\cdot,\cdot)$ is the *Hessian* of $J_\nu$ at $\Lambda$. The continuity of $\delta^2 J_\nu$ can be established by using the previous argumentation. In similar way, we can show that $J_\nu$ has continuous directional derivatives of any order, i.e. $J_\nu \in \mathcal{C}^k(\mathcal{L}_\nu^\Gamma)$ for any $k$. Finally, it remains to be shown that $J_\nu$ is strictly convex on the open set $\mathcal{L}_\nu^\Gamma$. Since $J_\nu \in \mathcal{C}^\infty(\mathcal{L}_\nu^\Gamma)$, it is sufficient to show that $\mathcal{H}_\Lambda(\delta\Lambda,\delta\Lambda) \geq 0$ for each $\delta\Lambda \in \mathrm{Range}\ \Gamma$ and equality holds if and only if $\delta\Lambda = 0$. Since $\nu > 0$ and the integrands in (48) are positive semidefinite when $\delta\Lambda_1 = \delta\Lambda_2$, we have $\mathcal{H}_\Lambda(\delta\Lambda,\delta\Lambda) \geq 0$. If $\mathcal{H}_\Lambda(\delta\Lambda,\delta\Lambda) = 0$, then $G^*\delta\Lambda G \equiv 0$ namely $\delta\Lambda \in [\mathrm{Range}\ \Gamma]^\perp$ (see [31, Section III]). Since $\delta\Lambda \in \mathrm{Range}\ \Gamma$, it follows that $\delta\Lambda = 0$. In conclusion, the Hessian is positive definite and the dual functional is strictly convex on $\mathcal{L}_\nu^\Gamma$. ■

In view of Theorem 5.1, the dual problem $\min_\Lambda \left\{ J_\nu(\Lambda) \mid \Lambda \in \mathcal{L}_\nu^\Gamma \right\}$ admits at most one solution $\Lambda^\circ$. Since $\mathcal{L}_\nu^\Gamma$ is an open set, such a $\Lambda^\circ$ (if it does exist) annihilates the first directional derivative (45) for each $\delta\Lambda \in \mathcal{Q}_n$

$$\left\langle I - \int [G(\Psi^{-\frac{1}{\nu}} + \frac{1}{\nu} G^*\Lambda^\circ G)^{-\nu} G^*, \delta\Lambda \right\rangle = 0 \ \ \forall \delta\Lambda \in \mathcal{Q}_n \tag{49}$$

or, equivalently,

$$I = \int G(\Psi^{-\frac{1}{\nu}} + \frac{1}{\nu} G^*\Lambda^\circ G)^{-\nu} G^* = \int G\Phi_\nu(\Lambda^\circ) G^*. \tag{50}$$

This means that $\Phi_\nu(\Lambda^\circ) \in \mathbb{S}_+^m(\mathbb{T})$ satisfies the constraint in (11) and $\Phi_\nu(\Lambda^\circ)$ is therefore the unique solution to Problem 2.

The next step concerns the existence issue for the dual problem. Although the existence question is quite delicate, since set $\mathcal{L}_\nu^\Gamma$ is open and unbounded, we will show that a $\Lambda^\circ$ minimizing $J_\nu$ over $\mathcal{L}_\nu^\Gamma$ does exist.

*Theorem 5.2:* Let $\nu \in \mathbb{N}_+ \setminus \{1\}$, then the dual functional $J_\nu$ has a unique minimum point in $\mathcal{L}_\nu^\Gamma$.

*Proof:* Since the solution of the dual problem (if it does exist) is unique, we only need to show that $J_\nu$ takes a minimum value on $\mathcal{L}_\nu^\Gamma$. First of all, note that $J_\nu$ is continuous on $\mathcal{L}_\nu^\Gamma$, see Theorem 5.1. Secondly, we show that $\mathrm{tr}[\Lambda]$ is bounded from below on $\mathcal{L}_\nu^\Gamma$. Since Problem 2 is feasible, there exists $\Phi_I \in \mathbb{S}_+^m(\mathbb{T})$ such that $\int G\Phi_I G^* = I$. Thus,

$$\mathrm{tr}[\Lambda] = \mathrm{tr}\left[\int G\Phi_I G^*\Lambda\right] = \mathrm{tr}\left[\int G^*\Lambda G\Phi_I\right]. \tag{51}$$

Defining $\alpha = -\nu \operatorname{tr} \int \Psi^{-\frac{1}{\nu}} \Phi_I$, we obtain

$$\operatorname{tr}[\Lambda] = \nu \operatorname{tr} \left[ \int (\Psi^{-\frac{1}{\nu}} + \frac{1}{\nu} G^* \Lambda G) \Phi_I \right] + \alpha. \tag{52}$$

Since $\Psi^{-\frac{1}{\nu}} + \frac{1}{\nu} G^* \Lambda G$ is positive definite on $\mathbb{T}$ for $\Lambda \in \mathcal{L}_\nu^\Gamma$, there exists a right spectral factor $\Delta$ such that $\Psi^{-\frac{1}{\nu}} + \frac{1}{\nu} G^* \Lambda G = \Delta^* \Delta$. Moreover, $\Phi_I$ is a coercive spectrum, namely there exists a constant $\mu > 0$ such that $\Phi_I(\mathrm{e}^{j\vartheta}) \geq \mu I$, $\forall\, \mathrm{e}^{j\vartheta} \in \mathbb{T}$. Starting from the fact that the trace and the integral are monotonic functions, we get

$$\begin{aligned} \operatorname{tr}[\Lambda] &= \nu \operatorname{tr} \left[ \int \Delta \Phi_I \Delta^* \right] + \alpha \geq \nu \mu \operatorname{tr} \left[ \int \Delta \Delta^* \right] + \alpha \\ &= \nu \mu \operatorname{tr} \left[ \int \Psi^{-\frac{1}{\nu}} + \frac{1}{\nu} G^* \Lambda G \right] + \alpha > \alpha \end{aligned} \tag{53}$$

where we have used the fact that $\operatorname{tr} \int \Psi^{\frac{1}{\nu}} + \frac{1}{\nu} G^* \Lambda G > 0$ when $\Lambda \in \mathcal{L}_\nu^\Gamma$. Finally, notice that $J_\nu(0) = -\frac{\nu}{1-\nu} \int \operatorname{tr}(\Psi^{\frac{\nu-1}{\nu}})$. Accordingly, we can restrict the search of a minimum point to the set $\left\{ \Lambda \in \mathcal{L}_\nu^\Gamma \mid J_\nu(\Lambda) \leq J_\nu(0) \right\}$. We now show that this set is compact. Accordingly, the existence of the solution to the dual problem follows from the Weierstrass' Theorem. To prove the *compactness* of the set, it is sufficient to show that:

1) $\lim\limits_{\Lambda \to \partial \mathcal{L}_\nu^\Gamma} J_\nu(\Lambda) = +\infty$;
2) $\lim\limits_{\|\Lambda\| \to \infty} J_\nu(\Lambda) = +\infty$.

1) Firstly, $R_\Lambda(z) := \Psi^{-\frac{1}{\nu}}(z) + \frac{1}{\nu} G(z)^* \Lambda G(z)$ is a rational matrix function, thus $R_\Lambda(z)^{1-\nu}$ is rational as well. Observe that $\partial \mathcal{L}_\nu^\Gamma$ is the set of $\Lambda \in \operatorname{Range} \Gamma$ such that $R_\Lambda(\mathrm{e}^{j\vartheta}) \geq 0$ on $\mathbb{T}$ and there exists $\vartheta$ such that $R_\Lambda(\mathrm{e}^{j\vartheta})$ is singular. Thus, for $\Lambda \to \partial \mathcal{L}_\nu^\Gamma$ all the eigenvalues of $R_\Lambda(z)^{-1}$ are positive on $\mathbb{T}$ and at least one of them has a pole tending to the unit circle. Since $1-\nu < -1$, then also $R_\Lambda(z)^{1-\nu}$ has at least one eigenvalue with a pole tending to $\mathbb{T}$. Accordingly, $\operatorname{tr}[\int R^{1-\nu}] \to \infty$ as $\Lambda \to \partial \mathcal{L}_\nu^\Gamma$. In view of (53), we conclude that $J_\nu(\Lambda) = -\frac{\nu}{1-\nu} \operatorname{tr} \left[ \int R^{1-\nu} \right] + \operatorname{tr}[\Lambda] \to \infty$ as $\Lambda \to \partial \mathcal{L}_\nu^\Gamma$.

2) Consider a sequence $\{\Lambda_k\}_{k \in \mathbb{N}} \in \mathcal{L}_\nu^\Gamma$, such that

$$\lim_{k \to \infty} \|\Lambda_k\| = \infty. \tag{54}$$

Let $\Lambda_k^0 = \frac{\Lambda_k}{\|\Lambda_k\|}$. Since $\mathcal{L}_\nu^\Gamma$ is convex and $0 \in \mathcal{L}_\nu^\Gamma$, if $\Lambda \in \mathcal{L}_\nu^\Gamma$ then $\xi \Lambda \in \mathcal{L}_\nu^\Gamma \quad \forall\, \xi \in [0,1]$. Therefore $\Lambda_k^0 \in \mathcal{L}_\nu^\Gamma$ for $k$ sufficiently large. Let $\eta := \liminf \operatorname{tr}[\Lambda_k^0]$. In view of (53),

$$\operatorname{tr}[\Lambda_k^0] = \frac{1}{\|\Lambda_k\|} \operatorname{tr}[\Lambda_k] > \frac{1}{\|\Lambda_k\|} \alpha \to 0, \tag{55}$$

for $k \to \infty$, so $\eta \geq 0$. Thus, there exists a subsequence of $\{\Lambda_k^0\}$ such that the limit of its trace is equal to $\eta$. Moreover, this subsequence remains on the surface of the unit ball $\partial\mathcal{B} = \left\{\Lambda = \Lambda^T \mid \|\Lambda\| = 1\right\}$ which is compact. Accordingly, it has a subsequence $\{\Lambda_{k_i}^0\}$ converging in $\partial\mathcal{B}$. Let $\Lambda^\infty \in \partial\mathcal{B}$ be its limit, thus $\lim_{i\to\infty} \mathrm{tr}[\Lambda_{k_i}^0] = \mathrm{tr}[\Lambda^\infty] = \eta$. We now prove that $\Lambda^\infty \in \mathcal{L}_\nu^\Gamma$. First of all, note that $\Lambda^\infty$ is the limit of a sequence in the finite dimensional linear space $\mathrm{Range}\,\Gamma$, hence $\Lambda^\infty \in \mathrm{Range}\,\Gamma$. It remains to be shown that $\Psi^{-\frac{1}{\nu}} + \frac{1}{\nu} G^* \Lambda^\infty G$ is positive definite on $\mathbb{T}$. Consider the unnormalized sequence $\{\Lambda_{k_i}\} \in \mathcal{L}_\nu^\Gamma$: We have that $\Psi^{-\frac{1}{\nu}} + \frac{1}{\nu} G^* \Lambda_{k_i} G > 0$ on $\mathbb{T}$ so that $\frac{1}{\|\Lambda_{k_i}\|}\Psi^{-\frac{1}{\nu}} + \frac{1}{\nu} G^* \Lambda_{k_i}^0 G$ is also positive definite on $\mathbb{T}$ for each $i$. Taking the limit for $i \to \infty$, we get that $G^*\Lambda^\infty G$ is positive semidefinite on $\mathbb{T}$ so that $\Psi^{-\frac{1}{\nu}} + \frac{1}{\nu} G^* \Lambda^\infty G > 0$ on $\mathbb{T}$. Hence, $\Lambda^\infty \in \mathcal{L}_\nu^\Gamma$. Since Problem 2 is feasible, there exists $\Phi_I \in \mathbb{S}_+^m(\mathbb{T})$ such that $I = \int G\Phi_I G^*$, accordingly

$$\eta = \mathrm{tr}[\Lambda^\infty] = \mathrm{tr} \int G\Phi_I G^* \Lambda^\infty = \mathrm{tr} \int \Phi_I^{\frac{1}{2}} G^* \Lambda^\infty G \Phi_I^{\frac{1}{2}}. \tag{56}$$

Moreover, $G^*\Lambda^\infty G$ is not identically equal to zero. In fact, if $G^*\Lambda^\infty G \equiv 0$, then $\Lambda^\infty \in [\mathrm{Range}\,\Gamma]^\perp$ and $\Lambda^\infty \neq 0$ since it belongs to the surface of the unit ball. This is a contradiction because $\Lambda^\infty \in \mathrm{Range}\,\Gamma$. Thus, $G^*\Lambda^\infty G$ is not identically zero and $\eta > 0$. Finally, we have

$$\begin{aligned}
&\lim_{k\to\infty} J_\nu(\Lambda_k) \\
&\quad = \lim_{k\to\infty} -\frac{\nu}{1-\nu}\,\mathrm{tr}\left[\int (\Psi^{-\frac{1}{\nu}} + \frac{1}{\nu} G^*\Lambda_k G)^{1-\nu}\right] + \mathrm{tr}[\Lambda_k] \\
&\quad \geq \lim_{k\to\infty} \|\Lambda_k\|\,\mathrm{tr}[\Lambda_k^0] = \eta \lim_{k\to\infty} \|\Lambda_k\| = \infty.
\end{aligned} \tag{57}$$

■

It remains to deal with the case $\nu \in \mathbb{Z}$ such that $\nu < 0$. In this situation, the dual problem may not have solution: The minimum point for $J_\nu(\Lambda)$ may lie on $\partial\mathcal{L}_\nu^\Gamma$, since $J_\nu$ takes finite values on the boundary of $\mathcal{L}_\nu^\Gamma$.

## VI. Computation of $\Lambda^\circ$

We showed that the dual problem always admits a unique solution $\Lambda^\circ$ on $\mathcal{L}_\nu^\Gamma$ for $\nu \in \mathbb{N}_+$. In order to find $\Lambda^\circ$, we exploit the following matricial *Newton* algorithm with backtracking stage proposed in [31]:

1) Set $\Lambda_0 = I \in \mathcal{L}_\nu^\Gamma$;

2) At each iteration, compute the *Newton* step $\Delta_{\Lambda_i}$ by solving the linear equation $\mathcal{H}_{\nu,\Lambda_i}(\Delta_{\Lambda_i},\cdot) = -\nabla J_{\nu,\Lambda_i}(\cdot)$ where, once fixed $\Lambda_i$, $\nabla J_{\nu,\Lambda_i}(\cdot)$ and $\mathcal{H}_{\nu,\Lambda_i}(\cdot,\cdot)$ must be understood as a linear and bilinear form of (45) and (48), respectively;
3) Set $t_i^0 = 1$ and let $t_i^{k+1} = t_i^k/2$ until both of the following conditions hold:

$$\Lambda_i + t_i^k \Delta_{\Lambda_i} \in \mathcal{L}_\nu^\Gamma \tag{58}$$

$$J_\nu(\Lambda_i + t_i^k \Delta_{\Lambda_i}) < J_\nu(\Lambda_i) + \alpha t_i^k \left\langle \nabla J_{\nu,\Lambda_i}, \Delta_{\Lambda_i} \right\rangle \tag{59}$$

with $0 < \alpha < 1/2$;
4) Set $\Lambda_{i+1} = \Lambda_i + t_i^k \Delta_{\Lambda_i}$;
5) Repeat steps 2, 3 and 4 until $\|\nabla J_{\nu,\Lambda_i}(\cdot)\| < \varepsilon$ where $\varepsilon$ is a tolerance threshold. Then set $\Lambda^\circ = \Lambda_i$.

The computation of the search direction $\Delta_{\Lambda_i}$ is the most delicate part of the procedure. The corresponding linear equation reduces to

$$\frac{1}{\nu}\sum_{l=1}^{\nu} \int G Q_{\Lambda_i}^l G^* \Delta_{\Lambda_i} G Q_{\Lambda_i}^{\nu+1-l} G^* = \int G Q_{\Lambda_i}^\nu G^* - I \tag{60}$$

where $Q_\Lambda = W(I + G_1^* \Lambda G_1)^{-1} W^*$. By similar argumentations used in [16, Proposition 8.1], it is possible to prove that there exists a unique solution $\Delta_{\Lambda_i} \in \operatorname{Range} \Gamma$ to (60). Accordingly, we can easily compute $\Delta_{\Lambda_i}$ in this way:

1) Compute

$$Y = \int G Q_{\Lambda_i}^\nu G^* - I; \tag{61}$$

2) Compute a basis $\{\Sigma_1 \dots \Sigma_M\}$ for $\operatorname{Range} \Gamma$ from (7) and for each $\Sigma_k$, $k = 1 \dots M$, compute

$$Y_k = \frac{1}{\nu}\sum_{l=1}^{\nu} \int G Q_{\Lambda_i}^l G^* \Sigma_k G Q_{\Lambda_i}^{\nu+1-l} G^*; \tag{62}$$

3) Find $\{\alpha_k\}$ such that $Y = \sum_k \alpha_k Y_k$. Then set $\Delta_{\Lambda_i} = \sum_k \alpha_k \Sigma_k$.

Concerning the evaluation of the integrals in (59), (61) and (62), a sensible and efficient method based on spectral factorization techniques may be employed. For further details, including the checking of condition (58), we refer to Section VI in [31].

Finally, it is possible to prove that:

1) $J_\nu(\cdot) \in \mathcal{C}^\infty(\mathcal{L}_\nu^\Gamma)$ is strongly convex on the sublevel set $\mathcal{K} = \left\{\Lambda \in \mathcal{L}_\nu^\Gamma \mid J_\nu(\Lambda) \le J_\nu(\Lambda_0)\right\}$;
2) The *Hessian* is *Lipschitz* continuous in $\mathcal{K}$.

The proof follows the ones in [31, Section VII] and [17, Section VI-C] faithfully. These properties allow us to conclude that the proposed *Newton* algorithm globally converges, [32, Chapter 9]. In particular the rate of convergence is quadratic during the last stage. In this way, the solution to Problem 2 may be efficiently computed.

## VII. Simulations results - Part I

In order to test the features of the family of solutions $\Phi_\nu$ with $\nu \in \mathbb{N}_+$, we take into account the following comparison procedure:

1) Choose a zero mean stationary process $y = \{y_k; k \in \mathbb{Z}\}$ with spectral density $\Omega \in \mathbb{S}_+^m(\mathbb{T})$;
2) Design a filters bank $G(z)$ as in (1);
3) Set $\Psi = \int \Omega$ ($\Psi$ is constant and equal to the zeroth moment of $\Omega$)
4) Set $\hat{\Sigma} = \Sigma \in \mathrm{Range}\, \Gamma \cap \mathcal{Q}_{n,+}$, i.e. the corresponding spectrum approximation problem is feasible;
5) Solve Problem 2 (with $\mathcal{S}_\nu$) by means of the proposed algorithm with the chosen $\Psi$ and $\hat{\Sigma}^{-\frac{1}{2}} G(z)$ as filters bank.

In the above comparison procedure we assume to know $\Sigma$ and $\int \Omega$. In this way, we avoid the approximation errors introduced by the estimation of $\Sigma$ and $\int \Omega$ from the finite-length data $y_1 \dots y_N$. Concerning the design of the filter, its role consists in providing the interpolation conditions for the solution to the spectrum approximation problem. More specifically, a higher resolution can be attained by selecting poles in the proximity of the unit circle, with arguments in the range of frequency of interest, [14].

### A. Scalar case

We start by taking into account Example described in [31, Section VIII-B] (the unique difference is that we assume to know $\Sigma$ and $\int \Omega$). Consider the following ARMA process:

$$\begin{aligned} y(t) &= 0.5y(t-1) - 0.42y(t-2) + 0.602y(t-3) \\ &- 0.0425y(t-4) + 0.1192y(t-5) \\ &+ e(t) + 1.1e(t-1) + 0.08e(t-2) - 0.15e(t-3) \end{aligned} \tag{63}$$

where $e$ is a zero-mean *Gaussian* white noise with unit variance. In Figure 1, the spectral density $\Omega \in \mathbb{S}_+^1(\mathbb{T})$ of the ARMA process is depicted (gray line). $G(z)$ is structured according to the

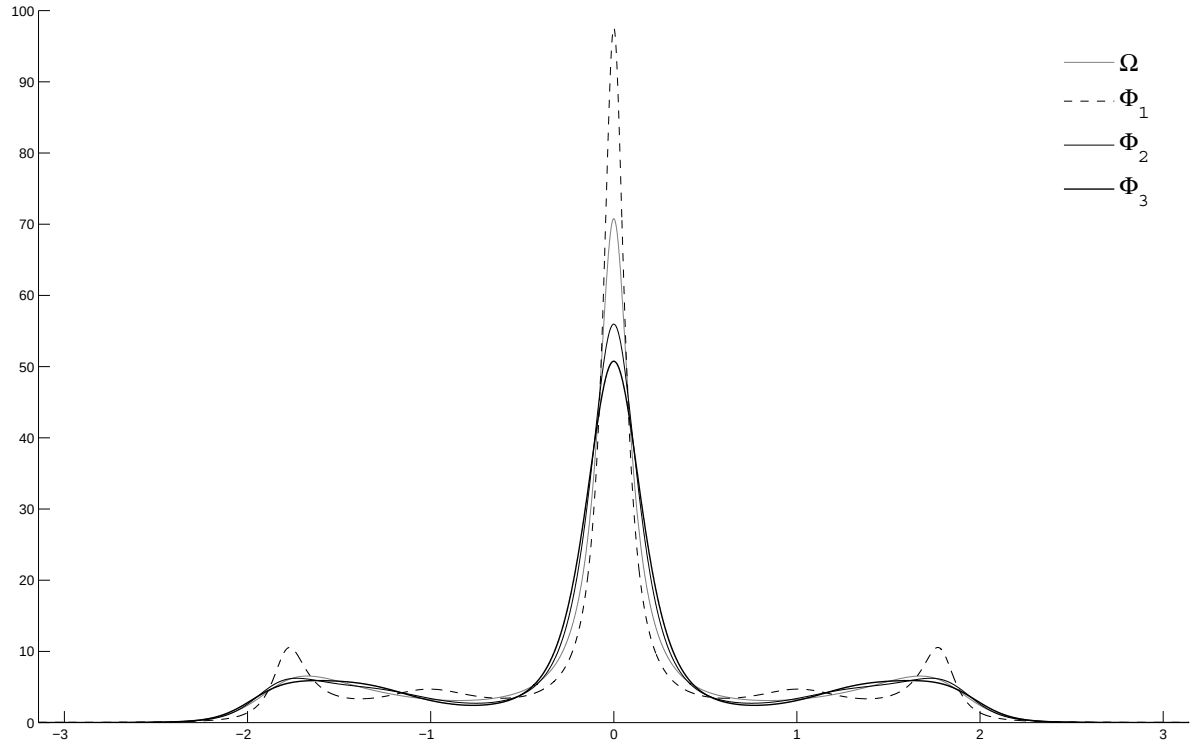


Fig. 1. Approximation of an ARMA $(6, 4)$ spectral density.

covariance extension setting (4) with $6$ covariance lags (i.e. $n = 6$). In Figure 1 the different solutions obtained by fixing $\nu = 1$, dashed line, $\nu = 2$, solid line, and $\nu = 3$, thick line, are shown. The solution obtained by minimizing the multivariate *Itakura-Saito* distance ($\nu = 1$) is characterized by peaks which are taller than these in $\Omega$. In fact, this solution seems the most adequate for detecting spectral lines, see example of Section VII-A in [17]. On the contrary, the peaks are reduced by increasing $\nu$. Note that, the solutions with $\nu = 2$ and $\nu = 3$ are closer to $\Omega$ than the one with $\nu = 1$.

As second example we consider the scalar *bandpass* random process with spectral density $\Omega$ depicted in Figure 2 (gray curve). The *cutoff* frequencies are $\vartheta_1 = 0.89$ and $\vartheta_2 = 2.46$. Moreover, $\Omega(\mathrm{e}^{j\vartheta}) \geq 2 \cdot 10^{-3}$ in the *stopband*, accordingly $\Omega \in \mathbb{S}_+^1(\mathbb{T})$. Matrix $B$ is a column of ones. Matrix $A$ is chosen as a block-diagonal matrix with one eigenvalue equal to zero and eight eigenvalues equispaced on the circle of radius $0.8$

$$\pm\, 0.8,\ 0.8\mathrm{e}^{\pm j\frac{\pi}{4}},\ 0.8\mathrm{e}^{\pm j\frac{\pi}{2}},\ 0.8\mathrm{e}^{\pm j\frac{3}{4}\pi}. \tag{64}$$

Here, $\Psi = \int \Omega \simeq 1.5284$. Figure 2 also shows the obtained solutions. The one with $\nu = 1$ turns

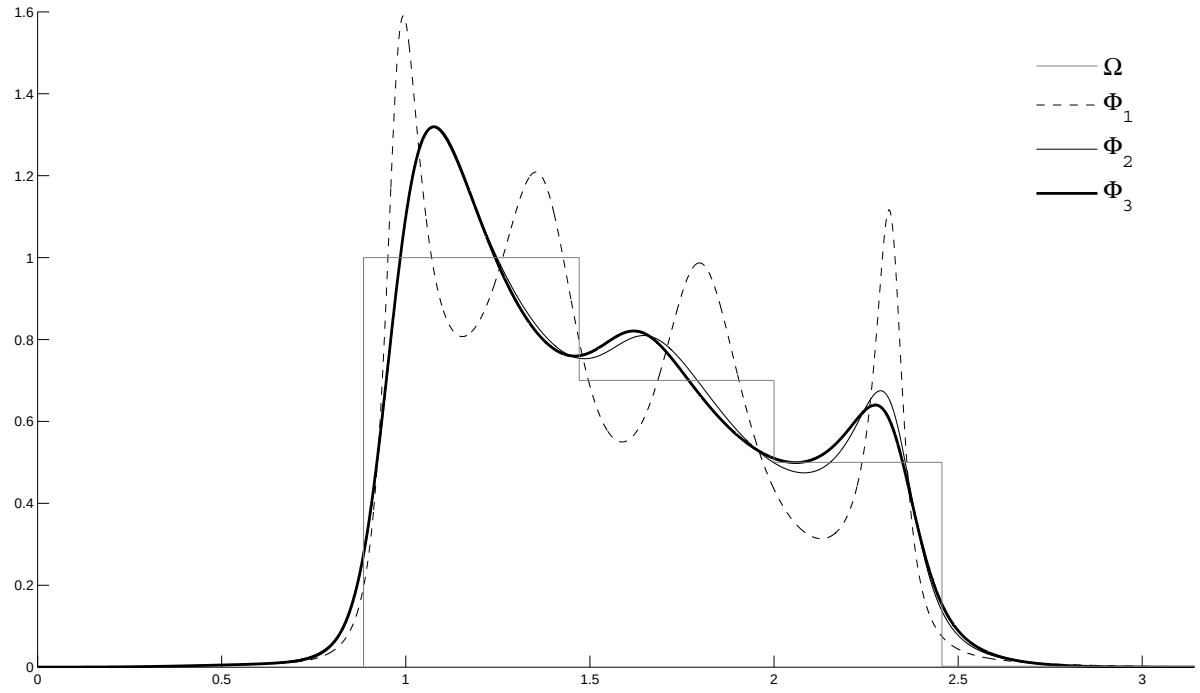


Fig. 2. Approximation of the spectral density of a scalar *bandpass* random process.

out inadequate. The solutions with $\nu = 2$ and $\nu = 3$ are, instead, similar and closer to $\Omega$.

### *B. Multivariate case*

We consider a bivariate *bandpass* random process with spectral density $\Omega$ plotted in Figure 3 (gray curve). Here, the *cutoff* frequencies are $\vartheta_1 = 0.42$ and $\vartheta_2 = 1.94$, and $\Omega(\mathrm{e}^{j\vartheta}) \geq 2 \cdot 10^{-3} I$ in the whole range of frequencies. The constant *prior* is

$$\Psi = \int \Omega \simeq \begin{pmatrix} 0.9313 & 0.3314 \\ 0.3314 & 0.5128 \end{pmatrix}. \tag{65}$$

The matrix $A$ of the filters bank has one eigenvalue equal to zero, two eigenvalues in $\pm 0.8$ and three pairs of complex eigenvalues closer to the passband $0.8\mathrm{e}^{\pm j0.4}$, $0.8\mathrm{e}^{\pm j1.2}$, $0.8\mathrm{e}^{\pm j2}$. The solutions for $\nu = 1$ (dashed line) $\nu = 2$ (solid line) and $\nu = 3$ (thick line) are shown in Figure 3. It is apparent that the solution for $\nu = 2$ is the most appropriate.

## VIII. Structured covariance estimation problem

As mentioned in Section II-A, we only have a *prior* $\Psi$ and a finite-length data $y_1 \dots y_N$ in the THREE-like spectral estimation procedure. Moreover, $\Phi_\nu$ represents a family of estimates of

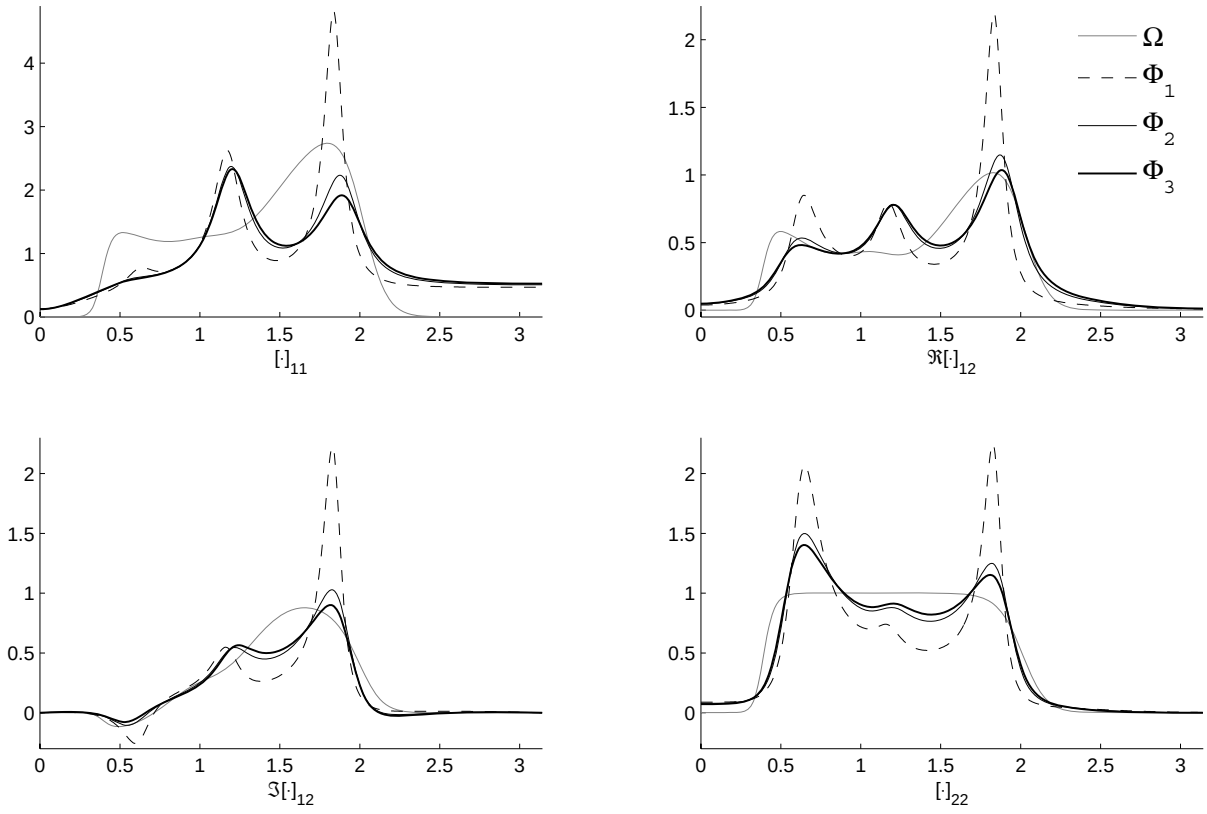


Fig. 3. Approximation of the spectral density of a bivariate *bandpass* random process.

$\Omega$ and we showed how to compute it starting from $\Psi$ and $\hat{\Sigma} \in \mathrm{Range}\,\Gamma \cap \mathcal{Q}_{n,+}$. Accordingly, it remains to find $\hat{\Sigma}$ from $y_1 \dots y_N$. To deal with this issue, we consider Problem 1 which can be viewed as the static version of Problem 2. Indeed, in both problems minimization of a divergence index, with respect to the first argument, is performed on the intersection among a vector space and an open cone. In this section, we briefly show that it is also possible to find a family of solutions to the structured covariance estimation problem.

The Beta matrix divergence (family) among two covariance matrices $P, Q \in \mathcal{Q}_{n,+}$ with $\beta \in \mathbb{R} \setminus \{0, 1\}$ is defined as

$$\mathcal{D}_\beta(P\|Q) := \mathrm{tr}[\frac{1}{\beta - 1}(P^\beta - PQ^{\beta-1}) - \frac{1}{\beta}(P^\beta - Q^\beta)]. \tag{66}$$

In fact, $\mathcal{D}_\beta(P\|Q)$ is the Beta divergence $\mathcal{S}_\beta(\Phi\|\Psi)$ among the two constant spectral densities $\Phi(\mathrm{e}^{j\vartheta}) \equiv P$ and $\Psi(\mathrm{e}^{j\vartheta}) \equiv Q$. Since $\mathcal{D}_\beta$ is a special case of $\mathcal{S}_\beta$, it is strictly convex with respect

to the first argument. Moreover, it is a continuous function of real variable $\beta \in \mathbb{R}$ with

$$\begin{aligned} \lim_{\beta \to 0} \mathcal{D}_\beta(P\|Q) &= \mathcal{D}_{\mathrm{B}}(P\|Q) \\ \lim_{\beta \to 1} \mathcal{D}_\beta(P\|Q) &= \mathcal{D}_{\mathrm{KL}}(P\|Q) \end{aligned} \tag{67}$$

where $\mathcal{D}_{\mathrm{B}} = 2\mathcal{D}_{\mathrm{I}}$ (see (10)) is the *Burg matrix* divergence, and

$$\mathcal{D}_{\mathrm{KL}}(P\|Q) := \mathrm{tr}[P(\log(P) - \log(Q)) - P + Q] \tag{68}$$

is the extension of the *Umegaki-von Neumann*'s relative entropy, [27], to non equal-trace matrices.

Take into account Problem 1 with $\mathcal{D}_\nu(P\|\hat{\Sigma}_C) := \mathcal{D}_\beta(P\|\hat{\Sigma}_C)$ such that $\beta = -\frac{1}{\nu} + 1$ and $\nu \in \mathbb{N}_+$. In [18], the existence and uniqueness of the solution to the problem for $\nu = 1$ has been showed. Moreover, the form of the optimal solution is $P_{\mathrm{B}}(\Delta) = [\hat{\Sigma}_C^{-1} + V^\star(\Delta)]^{-1}$, where $V^\star(\Delta) := \Pi_B^\perp \Delta \Pi_B^\perp - A^T \Pi_B^\perp \Delta \Pi_B^\perp A$ is the adjoint operator of the linear map $V$ defined in (8) and $\Delta \in \mathcal{Q}_n$ is the *Lagrange* multiplier. Consider now Problem 1 with $\nu \in \mathbb{N}_+ \setminus \{1\}$. The corresponding *Lagrange* functional is

$$\begin{aligned} L_\nu(P, \Delta) &:= \mathcal{D}_\nu(P\|\hat{\Sigma}_C) + \frac{\nu}{1-\nu} \mathrm{tr}(\hat{\Sigma}_C^{\frac{\nu-1}{\nu}}) + \langle V(P), \Delta \rangle \\ &= \mathcal{D}_\nu(P\|\hat{\Sigma}_C) + \frac{\nu}{1-\nu} \mathrm{tr}(\hat{\Sigma}_C^{\frac{\nu-1}{\nu}}) + \langle P, V^\star(\Delta) \rangle . \end{aligned} \tag{69}$$

Since $L_\nu(P, \Delta + \bar{\Delta}) = L_\nu(P, \Delta)\ \forall \bar{\Delta} \in \ker(V^\star)$, we can assume that $\Delta \in [\ker(V^\star)]^\perp$. Moreover, $L_\nu(\cdot, \Delta)$ is strictly convex over $\mathcal{Q}_{n,+}$. Thus, the unique minimum point of $L_\nu(\cdot, \Delta)$, which is given by annihilating the first directional derivative of $L_\nu(\cdot, \Delta)$, is

$$P_\nu(\Delta) := [\hat{\Sigma}_C^{-\frac{1}{\nu}} + \frac{1}{\nu} V^\star(\Delta)]^{-\nu}. \tag{70}$$

Since $P_\nu(\Delta) \in \mathcal{Q}_{n,+}$, the set of the admissible *Lagrange* multipliers is

$$\mathcal{L}_\nu := \left\{ \Delta \in \mathcal{Q}_n \mid \hat{\Sigma}_C^{-\frac{1}{\nu}} + \frac{1}{\nu} V^\star(\Delta) > 0 \right\} \cap [\ker(V^\star)]^\perp \tag{71}$$

which is an open and bounded set (the proof is similar to the one of Proposition 5.1 in [18]). Then, the dual problem is

$$\Delta^\circ = \operatorname*{argmin}_{\Delta \in \mathcal{L}_\nu} J_\nu(\Delta) \tag{72}$$

where

$$\begin{aligned} J_\nu(\Delta) &:= -\inf_P L_\nu(P, \Delta) \\ &= \frac{\nu}{\nu - 1} \mathrm{tr}[\hat{\Sigma}_C^{-\frac{1}{\nu}} + \frac{1}{\nu} V^\star(\Delta)]^{1-\nu}. \end{aligned} \tag{73}$$

Note that $J_\nu(0) = \frac{\nu}{\nu-1}\operatorname{tr}(\hat{\Sigma}_C^{\frac{\nu-1}{\nu}})$. Accordingly, we can restrict the search of a minimum point to the set $\mathcal{L}^\star := \left\{\Delta \in \mathcal{L}_\nu^\Gamma \mid J_\nu(\Delta) \leq J_\nu(0)\right\} \subset \mathcal{L}_\nu^\Gamma$ which is bounded. Following the same lines in [18], it is possible to prove that $J_\nu \in \mathcal{C}^\infty(\mathcal{L}_\nu)$ is strictly convex on $\mathcal{L}_\nu$ and $\lim_{\Delta\to\partial\mathcal{L}_\nu} J_\nu(\Delta) = +\infty$ (the limit diverges because the exponent in (73) is negative). Thus, $\mathcal{L}^\star$ is a compact set (i.e. closed and bounded) and $J_\nu$ admits a minimum point $\Delta^\circ$ over $\mathcal{L}^\star$ by the Weierstrass' Theorem. The uniqueness of $\Delta^\circ$ follows from the fact that $J_\nu$ is strictly convex over $\mathcal{L}_\nu^\Gamma$. Also in this case, a globally convergent matricial *Newton* algorithm for finding $\Delta^\circ$ may be employed. Therefore, once we computed $\Delta^\circ$ the solution to Problem 1 is given by $P_\nu(\Delta^\circ)$. Finally, the same analysis may be extended to $\mathcal{D}_{\mathrm{KL}}$. In this case, $P_{KL}(\Delta) = \mathrm{e}^{\log(\hat{\Sigma}_C) - V^\star(\Delta)}$.

To sum up, a family of solutions $P_\nu \in \operatorname{Range}\Gamma \cap \mathcal{Q}_{n,+}$ to the structured covariance estimation problem has been found. In this way, we have a complete tool to compute the family of estimates $\Phi_\nu$ of $\Omega$ starting from a *prior* $\Psi$ and a finite-length data $y_1 \dots y_N$: We compute $P_\nu$ from $y_1 \dots y_N$ and we then find $\Phi_\nu$ starting from $P_\nu$ and $\Psi$.

## IX. Simulation results - Part II

We consider the bivariate *bandpass* random process $y$ of Section VII-B and we take into account the following THREE-like spectral estimation procedure:

1) We start from a finite sequence $y_1 \dots y_N$ extracted from a realization of the process $y$;
2) Fix $G(z)$ as in Section VII-B;
3) Set $\Psi = \frac{1}{N}\sum_{k=1}^N y_k y_k^T$;
4) Feed the filters bank with the data sequence $y_1 \dots y_N$, collect the output data $x_1 \dots x_N$ and compute $\hat{\Sigma}_C = \frac{1}{N}\sum_{k=1}^N x_k x_k^T$;
5) Compute $P_\nu \in \operatorname{Range}\Gamma \cap \mathcal{Q}_{n,+}$ by solving Problem 1 (with $\mathcal{D}_\nu$), then set $\hat{\Sigma} = P_\nu$
6) Compute $\Phi_\nu$ by solving Problem 2 (with $\mathcal{S}_\nu$) by means of the proposed algorithm with the chosen $\Psi$ and $\hat{\Sigma}^{-\frac{1}{2}} G(z)$ as filters bank.

Note that, the *prior* $\Psi$ represents a coarse, low order, estimate of $\Omega$ obtained by the means of a standard and simple estimation method. Accordingly, $\Phi_\nu$ is a spectral density (with bounded *McMillan* degree) which is consistent with the interpolation constraint in (3) and is as close as possible to the initial estimate $\Psi$ according to the divergence index $\mathcal{S}_\nu$. In the above procedure, $\Psi$ has been chosen as the constant spectral density equal to the variance of the given data sequence.

In Figure 4, the obtained estimates with $N = 50$ (i.e. we have considered a short-length data) are depicted. Also in this case, the peaks of the estimates are reduced by increasing $\nu$. For the extracted sequence, the estimators for $\nu = 2$ and $\nu = 3$ appear to perform better than the one for $\nu = 1$. Finally, the same procedure can be applied to the other processes considered in Section

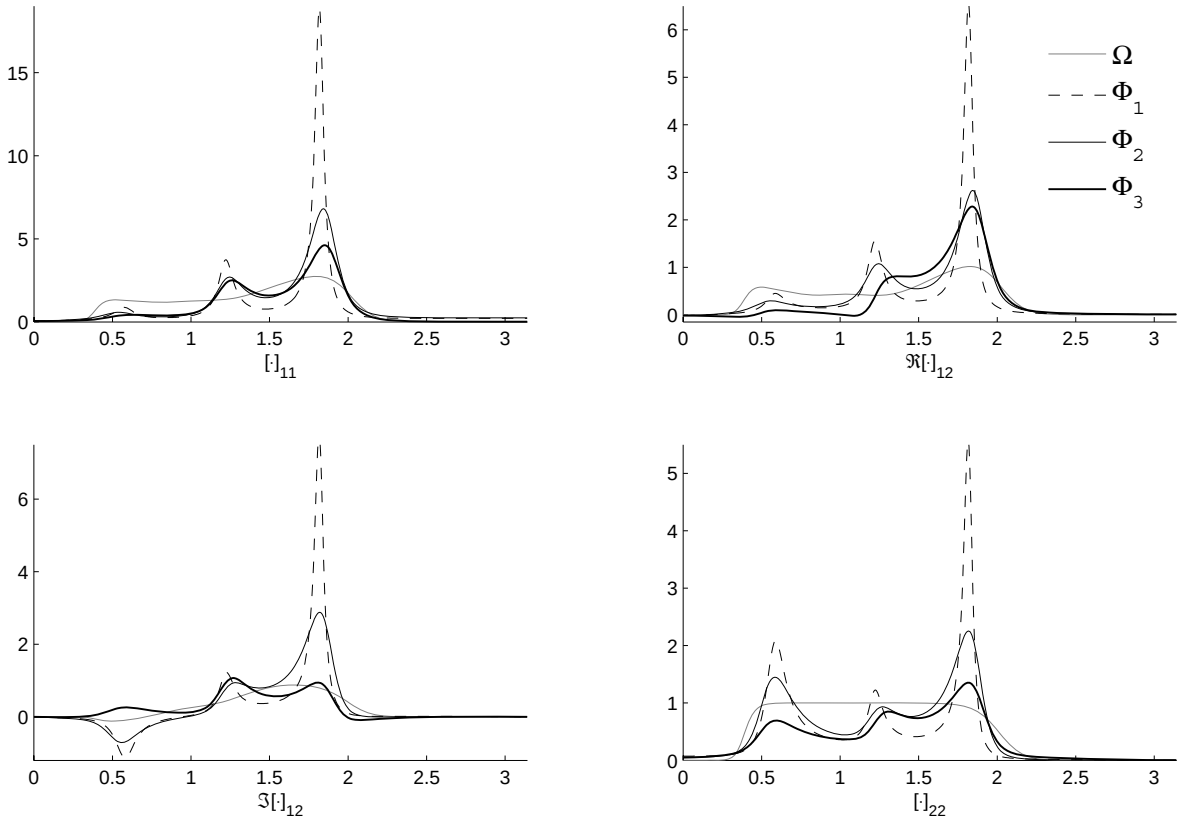


Fig. 4. Estimation of the spectral density of a bivariate *bandpass* random process.

VII.

## X. Conclusions

A multivariate Beta divergence family connecting the *Itakura-Saito* distance with the *Kullback-Leibler* divergence has been introduced. The corresponding solutions to the spectrum approximation problem are rational when a suitable parametrization of the parameter $\beta$ is employed. The solution with the smallest upper bound on the *McMillan* degree is associated to the *Itakura-Saito* distance. Moreover, the limit of this family tends to the solution corresponding to the *Kullback-Leibler* divergence. Then, similar results may be found for the structured covariance estimation

problem. Simulations suggest that the presented family of estimators provides a relevant tool in multivariate spectral estimation.

## ACKNOWLEDGMENTS

The author wish to thank Prof. A. Ferrante and Prof. M. Pavon for useful discussions about THREE-like spectral estimation approaches.

## APPENDIX

### A. On the exponentiation of positive definite matrices

We collect some technical result concerning the exponentiation of positive definite matrices to an arbitrary real number. We start by introducing the differential of the matrix exponential and the matrix logarithm (see [15]).

*Proposition A.1:* Given $Y \in \mathcal{Q}_n$, the differential of $Y \mapsto \mathrm{e}^Y$ in the direction $\Delta \in \mathcal{Q}_n$ is given by the linear map

$$M_Y : \Delta \mapsto \int_0^1 \mathrm{e}^{(1-\tau)Y} \Delta \mathrm{e}^{\tau Y} \mathrm{d}\tau. \tag{74}$$

*Proposition A.2:* Given $Y \in \mathcal{Q}_{n,+}$, the differential of $Y \mapsto \log(Y)$ in the direction $\Delta \in \mathcal{Q}_n$ is given by the linear map

$$N_Y : \Delta \mapsto \int_0^\infty (Y + tI)^{-1} \Delta (Y + tI)^{-1} \mathrm{d}t. \tag{75}$$

Let us consider now a positive definite matrix $X \in \mathcal{Q}_{n,+}$ and a real number $c$. The exponentiation of $X$ to $c$ may be rewritten in the following way

$$X^c = e^{c \log X}. \tag{76}$$

Accordingly, by applying the chain rule, the differential of $X \mapsto X^c$ in the direction $\Delta \in \mathcal{Q}_n$ is given by

$$M_{c \log X}(cN_X(\Delta)) = c \int_0^1 X^{c(1-\tau)} \int_0^\infty (X + tI)^{-1} \Delta (X + tI)^{-1} \mathrm{d}t X^{c\tau} \mathrm{d}\tau. \tag{77}$$

We summarize this result below.

*Proposition A.3:* The differential of $X \mapsto X^c$ in direction $\Delta \in \mathcal{Q}_n$ is given by the linear map

$$O_{X,c} : \Delta \mapsto c \int_0^1 X^{c(1-\tau)} \int_0^\infty (X + tI)^{-1} \Delta (X + tI)^{-1} \mathrm{d}t X^{c\tau} \mathrm{d}\tau. \tag{78}$$

*Corollary A.1:* The first variation of $X \mapsto \mathrm{tr}(X^c)$ in direction $\Delta \in \mathcal{Q}_n$ is

$$\delta(\mathrm{tr}(X^c); \Delta) = c\,\mathrm{tr}(X^{c-1}\Delta). \tag{79}$$

*Proof:* Since $X^{c\tau}$ and $(X + tI)$ commute, we get

$$\begin{aligned}
\delta(\mathrm{tr}(X^c); \Delta) &= \mathrm{tr}(O_{X,c}(\Delta)) \\
&= c\,\mathrm{tr}\left\{X^c \int_0^\infty (X+tI)^{-2}\mathrm{d}t\Delta\right\} \\
&= c\,\mathrm{tr}\left\{X^c X^{-1}\Delta\right\} = c\,\mathrm{tr}\left\{X^{c-1}\Delta\right\}.
\end{aligned}$$

■

## REFERENCES


[1] X. Jiang, L. Ning, and T. Georgiou, "Distances and riemannian metrics for multivariate spectral densities," *IEEE Trans. Autom. Control*, vol. 57, no. 7, pp. 1723–1735, Jul. 2012.

[2] T. Georgiou, "Realization of power spectra from partial covariance sequences," *IEEE Trans. Acoust., Speech Signal Processing*, vol. 35, no. 4, pp. 438–449, Apr. 1987.

[3] C. Byrnes, A. Lindquist, S. Gusev, and A. Matveev, "A complete parameterization of all positive rational extensions of a covariance sequence," *IEEE Trans. Autom. Control*, vol. 40, no. 11, pp. 1841–1857, Nov 1995.

[4] C. Byrnes, T. Georgiou, and A. Lindquist, "A generalized entropy criterion for Nevanlinna-Pick interpolation with degree constraint," *IEEE Trans. Autom. Control*, vol. 46, no. 5, pp. 822–839, May 2001.

[5] T. Georgiou, "Spectral analysis based on the state covariance: The maximum entropy spectrum and linear fractional parametrization," *IEEE Trans. Autom. Control*, vol. 47, no. 11, pp. 1811–1823, Nov. 2002.

[6] C. Byrnes and A. Lindquist, "On the partial stochastic realization problem," *IEEE Trans. Autom. Control*, vol. 42, no. 8, pp. 1049–1070, Aug 1997.

[7] C. Byrnes, S. Gusev, and A. Lindquist, "A convex optimization approach to the rational covariance extension problem," *SIAM J. Control and Optimization*, vol. 37, no. 1, pp. 211–229, Oct. 1998.

[8] C. Byrnes, T. Georgiou, A. Lindquist, and A. Megretski, "Generalized interpolation in H-infinity with a complexity constraint," *Trans. American Math. Society*, vol. 358, no. 3, pp. 965–987, Dec. 2004.

[9] T. Georgiou and A. Lindquist, "Remarks on control design with degree constraint," *IEEE Trans. Autom. Control*, vol. 51, no. 7, pp. 1150–1156, Jul. 2006.

[10] A. Nasiri Amini, E. Ebbini, and T. Georgiou, "Noninvasive estimation of tissue temperature via high-resolution spectral analysis techniques," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 2, pp. 221–228, Feb. 2005.

[11] C. Byrnes, P. Enqvist, and A. Lindquist, "Identifiability and well-posedness of shaping-filter parameterizations: A global analysis approach," *SIAM J. Control and Optimization*, vol. 41, no. 1, pp. 23–59, Mar. 2002.

[12] T. Georgiou and A. Lindquist, "A convex optimization approach to ARMA modeling," *IEEE Trans. Autom. Control*, vol. 53, no. 5, pp. 1108–1119, Jun. 2008.

[13] J. Karlsson and T. Georgiou, "Uncertainty Bounds for Spectral Estimation," *ArXiv e-prints*, 2012.

[14] C. Byrnes, T. Georgiou, and A. Lindquist, "A new approach to spectral estimation: A tunable high-resolution spectral estimator," *IEEE Trans. Signal Processing*, vol. 48, no. 11, pp. 3189–3205, Nov. 2000.

[15] T. Georgiou, "Relative entropy and the multivariable multidimensional moment problem," *IEEE Trans. Inform. Theory*, vol. 52, no. 3, pp. 1052–1066, Mar. 2006.

[16] A. Ferrante, M. Pavon, and F. Ramponi, "Hellinger versus Kullback-Leibler multivariable spectrum approximation," *IEEE Trans. Autom. Control*, vol. 53, no. 4, pp. 954–967, May 2008.

[17] A. Ferrante, C. Masiero, and M. Pavon, "Time and spectral domain relative entropy: A new approach to multivariate spectral estimation," *IEEE Trans. Autom. Control*, vol. 57, no. 10, pp. 2561–2575, Oct. 2012.

[18] A. Ferrante, M. Pavon, and M. Zorzi, "A maximum entropy enhancement for a family of high-resolution spectral estimators," *IEEE Trans. Autom. Control*, vol. 57, no. 2, pp. 318–329, Feb. 2012.

[19] M. Zorzi and A. Ferrante, "On the estimation of structured covariance matrices," *Automatica*, vol. 48, no. 9, pp. 2145–2151, Sep. 2012.

[20] A. Basu, I. Harris, N. Hjort, and M. Jones, "Robust and efficient estimation by minimising a density power divergence," *Biometrika*, vol. 85, no. 3, pp. 549–559, 1998.

[21] M. Mollah, N. Sultana, M. Minami, and S. Eguchi, "Robust extraction of local structures by the minimum of Beta-divergence method," *Neural Networks*, vol. 23, no. 2, pp. 226–238, 2010.

[22] M. Minami and S. Eguchi, "Robust blind source separation by Beta divergence," *Neural Computation*, vol. 14, no. 8, pp. 1859–1886, Aug. 2002.

[23] A. Cichocki, R. Zdunek, A. Phan, and S. Amari, *Nonnegative Matrix and Tensor Factorizations*. Chichester, U.K.: John Wiley & Sons, 2009.

[24] A. Cichocki, S. Cruces, and S. Amari, "Generalized Alpha-Beta divergences and their application to robust nonnegative matrix factorization," *Entropy*, vol. 13, no. 1, pp. 134–170, Jan. 2011.

[25] T. Georgiou, "The structure of state covariances and its relation to the power spectrum of the input," *IEEE Trans. Autom. Control*, vol. 47, no. 7, pp. 1056–1066, Jul. 2002.

[26] T. M. Cover and J. A. Thomas, *Information Theory*. New York: Wiley, 1991.

[27] M. Nielsen and I. Chuang, *Quantum Computation and Quantum Information*. Cambridge, U.K.: Cambridge University Press, 2004.

[28] A. Cichocki and S. Amari, "Families of Alpha- Beta- and Gamma- divergences: Flexible and robust measures of similarities," *Entropy*, vol. 12, no. 6, pp. 1532–1568, Jun. 2010.

[29] T. Georgiou and A. Lindquist, "Kullback-Leibler approximation of spectral density functions," *IEEE Trans. Inform. Theory*, vol. 49, no. 11, pp. 2910–2917, Nov. 2003.

[30] A. Blomqvist, A. Lindquist, and R. Nagamune, "Matrix-valued Nevanlinna-Pick interpolation with complexity constraint: An optimization approach," *IEEE Trans. Autom. Control*, vol. 48, no. 12, pp. 2172–2190, Dec. 2003.

[31] F. Ramponi, A. Ferrante, and M. Pavon, "A globally convergent matricial algorithm for multivariate spectral estimation," *IEEE Trans. Autom. Control*, vol. 54, no. 10, pp. 2376–2388, Oct. 2009.

[32] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge University Press, 2004.